THE PROCEEDINGS OF
THE PHYSICAL SOCIETY
Section B

VOL. 62, PART 1 | 1 January 1949 | No. 349 B

EDITORIAL

Beginning with this issue the *Proceedings* will appear in two Parts, "A" and "B". This change has been dictated by the increasing volume of important communications which has proved too great even for the larger amount of space available since monthly publication began in January 1948.

It is intended provisionally to publish in Part A papers dealing with subjects such as the quantum theory, statistical mechanics, nuclear physics and cosmic rays, atomic physics, molecules, spectra, theories of solids, liquids and gases, surface phenomena, growth and properties of crystals, crystal structure, luminescence, electrodynamics, heat and thermodynamics, and standards. Part B will include subjects such as acoustics (including ultrasonics), optical design, electron optics, colour, elasticity and other mechanical properties of solids and liquids, crystal structure analysis, magnetic materials, refrigeration and liquefaction, electric discharges, radio, geo- and ionosphere physics, astrophysics, and solar physics. It is obvious that a hard and fast dividing line cannot at present be drawn, but we feel sure that in a short time each Part will develop a character of its own, so that the reader will have little difficulty in deciding with which he wishes to be supplied.

With this new arrangement we hope to be able, not only to give improved service to Fellows and other subscribers, but also to effect that speedy publication of important scientific papers which is so vital to the development of physics.

G. INGLE FINCH,
President.

On the Theory of Aplanatic Aspheric Systems

By G. D. WASSERMANN* and E. WOLF†

H. H. Wills Physical Laboratory, University of Bristol

* Now at King's College, University of Durham

† Now at the Observatories, University of Cambridge

MS. received 21*st February* 1948; *read before Optical Group* 17*th December* 1948

ABSTRACT. Methods are described for the design of two aspheric surfaces for any given centred system, so as to achieve exact aplanatism. The practical application of the methods is illustrated by the design of a reflecting microscope.

---

§1. INTRODUCTION

THE design of aspherical systems raises a number of questions. Foremost among these is the problem of how to determine for any given system the aspheric profiles so as to obtain good definition over the whole of a finite field. To the approximation of third order (Seidel) optics this can be accomplished by the elegant method of plate diagram analysis (Burch 1942, Linfoot 1944). As a rule, however, higher aberrations must be taken into account. A reasonable way of doing this is to supplement the preliminary Seidel design by further analysis. General methods using differential corrections were described by Volosov (1947), but their form makes them difficult to handle. Frequently, however, it is sufficient to confine the further analysis to the attainment of either axial stigmatism or aplanatism.‡ The former has been discussed by Herzberger and Hoadley (1946) and by Wolf (1948). No corresponding analysis for the case of aplanatism appears to have been made. However, a simple method adequate for some systems was given by Linfoot (1943). Particular systems were considered by various authors; for instance, Schwarzschild (1905) and Chrétien (1922) derived exact equations for two-mirror aplanatic telescopes. Their methods, although valuable for the practical design of such aplanats, are not very suitable for application to more complex systems.§

The aim of the present paper is to provide methods for the design of two aspheric surfaces for any given centred system, so as to achieve exact aplanatism. We derive two simultaneous first-order differential equations for their profiles. These equations contain certain functions which characterize two-ray congruences in the spaces immediately before and after the two surfaces. These are easily obtained from a ray trace. Our differential equations permit of a rigorous solution of the problem and can be solved to any desired accuracy by standard numerical methods.

§2. EQUATIONS FOR THE DESIGN OF TWO SURFACES TO ENSURE APLANATISM

2.1. In this section we derive formulae for the design of two correcting surfaces so as to obtain simultaneously: (*a*) axial stigmatism, (*b*) exact satisfaction of the sine condition. Our results apply to any centred system in which the two aspheric surfaces are optical neighbours. The two surfaces in question may, however, be separated from the object or image points by any number of spherical or aspherical surfaces. We shall only be concerned with the final corrections of the system, and we assume all design data, save the profiles of the two correctors, to be known.

---

‡ We use the term aplanatism to imply axial stigmatism together with the satisfaction of the exact sine condition.

§ This is illustrated, for example, by the analysis of Bureau and Swings (1934) who applied Schwarzschild's methods to the design of two-mirror aplanats with both foci finite.

We introduce two sets of Cartesian coordinate systems (Figure 1) with origins at the poles O and O′ (assumed to be known) of the two correcting surfaces Σ and Σ', with their x axes along the axis of the system; and take $OO' = d$. Let us denote by σ and σ' the spaces immediately preceding Σ and following Σ', and by σ^* the space between them. Finally, let n, n^* and n' be the corresponding refractive indices.

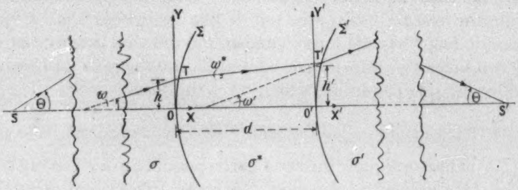


Figure 1.

The rays which proceed from the axial object point S are refracted (or reflected) at the successive surfaces and form normal rectilinear congruence Γ in the space σ. Let h denote the height at which a typical ray of Γ meets OY (if produced far enough) and let ω denote the angle which the rays make with the x axis (measured anticlockwise from the positive x direction to the direction along which the light advances). Γ can then be completely specified by two relations:

$$\omega = \omega(t); \qquad h = h(t), \qquad \ldots\ldots(2.1)$$

where t is a free parameter. If the axial object point S is at a finite distance we choose

$$t = \sin\Theta, \qquad \ldots\ldots(2.2)$$

where Θ is the angle which the corresponding ray in the object space makes with the x axis (Figure 1). If, on the other hand, the object is at infinity, we choose $t = H$ where H is the distance of the corresponding ray in the object space from the x axis (Figure 2(a)).

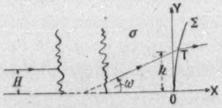


Figure 2 (a).

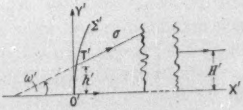


Figure 2 (b).

The ray specified by the parameter t will be referred to as the ray t and its intersection with Σ will be denoted by T. As the relations (2.1) can rarely, in practice, be obtained in analytical form, we form a table of ω and h based on a ray trace for different values of t. Some designers would probably favour polynomial approximations to $\omega(t)$ and $h(t)$ at this stage; but the merits of this expedient are doubtful and will be discussed later.

Similarly the rays which are traced backwards from the axial image point S′ are refracted (or reflected) in succession by all surfaces which follow Σ' and form a normal rectilinear congruence in the space σ'. Let Γ' be the congruence in σ' which is obtained by reversing the sense of every ray of the "back-traced" congruence, i.e. we regard the rays of Γ' as though they proceed from Σ' towards the axial image point (Figure 1). Γ' can then be specified by two relations:

$$\omega' = \omega'(t'); \qquad h' = h'(t'), \qquad \ldots\ldots(2.3)$$

where ω' and h' are now referred to axes at O′ and the parameter t' has a significance analogous to that of t (Figures 1 and 2(*b*)). The quantities ω' and h' are obtained for selected values of t' by tracing rays backwards from the axial image point S′ into the space σ'.

The conditions (*a*) and (*b*) for aplanatism require that the ray t of Γ be transformed by two successive refractions at Σ and Σ' into the ray t' of Γ' where t and t' are related by the sine condition, which takes the form

$$t/t' = \text{constant}. \qquad \ldots\ldots(2.4)$$

Equation (2.4) defines uniquely the ray t' corresponding to a given ray t, so that our problem can be regarded as involving only the free parameter t. To avoid cumbersome suffixes we shall use undashed and dashed symbols to refer to quantities associated with a pair of corresponding rays.

2.2. Let (x, y) and (x', y') be the coordinates, referred to axes at O and O′ respectively, of the points T and T′ (Figure 1). We have, by Snell's law, for the refraction at Σ

$$n(\cos\omega\, dx/dt + \sin\omega\, dy/dt) = n^*(\cos\omega^* dx/dt + \sin\omega^* dy/dt), \qquad \ldots\ldots(2.5)$$

where ω^* is the angle which TT′ makes with the axis.

We have from Figure 1

$$\sin\omega^* = R_y/R; \qquad \cos\omega^* = R_x/R, \qquad \ldots\ldots(2.6)$$

where

$$R_x = x' - x + d, \quad R_y = y' - y, \quad \text{and} \quad R^2 = R_x{}^2 + R_y{}^2. \qquad \ldots\ldots(2.7)$$

Also, from the figure,

$$y = h + x\tan\omega \qquad \ldots\ldots(2.8)$$

and

$$y' = h' + x'\tan\omega'. \qquad \ldots\ldots(2.9)$$

Substituting for $\cos\omega^*$ and $\sin\omega^*$ from (2.6) and for dy/dt from (2.8) into (2.5) it follows that

$$\frac{dx}{dt} = -\left[\frac{(n^*R_x - nR\cos\omega)}{(n^*R_y - nR\sin\omega)} + \tan\omega\right]^{-1}\left[\frac{dh}{dt} + x\frac{d}{dt}\tan\omega\right]. \qquad \ldots\ldots(2.10)$$

Similarly we obtain

$$\frac{dx'}{dt} = -\left[\frac{(n^*R_x - n'R\cos\omega')}{(n^*R_y - n'R\sin\omega')} + \tan\omega'\right]^{-1}\left[\frac{dh'}{dt} + x'\frac{d}{dt}\tan\omega'\right]. \qquad \ldots\ldots(2.11)$$

Equations (2.10) and (2.11) permit together with (2.7), (2.8) and (2.9) a complete computation of Σ and Σ'. For by means of (2.7), (2.8) and (2.9) we could eliminate

y and y' from (2.10) and (2.11) and thus obtain two simultaneous first-order differential equations of the type

$$dx/dt=f(x,x',t); \quad \text{and} \quad dx'/dt=g(x,x',t). \qquad \ldots\ldots(2.12)$$

These may be integrated numerically by standard methods. But as we require y and y' as well as x and x' for a range of values of the parameter t, it is preferable to avoid elimination and solve for the unknown quantities step by step. This will be discussed in more detail in the next section. Equation (2.12), subject to the boundary conditions $x=x'=0$ for $t=0$ show that if there exists a physical solution of our problem it will be unique.

2.3. If the correcting surface Σ is a mirror, the following rule should be applied: Put $n=-n^*=1$ and change R_x and R_y in (2.10) into $-R_x$ and $-R_y$. Similarly if Σ' is a mirror, put $n'=-n^*=1$ and change the sign of R_x and R_y in (2.11).

§ 3. METHODS OF COMPUTATION AND SOME APPLICATIONS

3.1. We shall first summarize the procedure for computing the two correcting surfaces.

(*a*) Rays are traced from the axial object point into the space σ for a set of selected values of the parameter t. For each value of t the quantities ω and h are determined. Following this we calculate the values of $d\tan\omega/dt$ and dh/dt by numerical differentiation or otherwise. In simple cases (e.g. when Σ is the first surface of the system) the differentiation may be performed analytically.

(*b*) For each value of t the corresponding value of t' is calculated from (2.4). The quantities ω' and h' are now determined by tracing rays backwards from the axial image point into the space σ'. The differential coefficients $d\tan\omega'/dt$ and dh'/dt are subsequently determined as above.

(*c*) Next, equations (2.10) and (2.11) are integrated numerically step by step, for every value of x and x' the corresponding values of y and y' are calculated at every step from (2.8) and (2.9). The integration is performed in two stages. The first few values can be obtained by the application of the methods of Runge and Kutta (Runge and König 1924, p. 286). The integration can then be continued by Adam's method (Whittaker and Robinson 1946, p. 363). In this way we obtain the two profiles in tabulated form.

Some opticians would probably employ throughout the design polynomial approximations, based on the tracings of a small number of rays. Although polynomial approximations form an important tool in preliminary design they are, in our opinion, not necessarily of advantage in the final computation of an aspheric system. Our doubts of the expediency of polynomial approximations arise first of all in connection with the high accuracy which is frequently required in the design of an aspheric surface. (In astrographic cameras for example, it is generally desirable to determine the profiles down to a fraction of a fringe.) On account of slow convergence (cf. Martin 1944, p. 108) this may necessitate the evaluation of a large number of coefficients. As a rule, however, it is difficult to obtain more than the first three or four coefficients in terms of the design parameters, so that the expansions have to be supplemented in particular cases. Moreover, if we base the expansion on the tracing of a small number of rays we encounter further difficulties as it is generally not possible to determine in advance the number of terms required for a satisfactory expansion (Herzberger and Hoadley 1946, p. 339).

In view of these considerations it is our opinion that polynomial approximations are not very suitable in the solution of this type of problem and should be replaced by tabulated functions based on tracings of dense fans of rays. Linfoot (1943, p. 494) also advocates this procedure.

3.2. To test the analysis we applied our methods to the design of an aplanatic reflecting microscope and compared the results with those obtained from the analytical solutions of Schwarzschild (1905) and Chrétien (1922) for two-mirror aplanats. Microscopes of this type were constructed and discussed by Burch (1947).

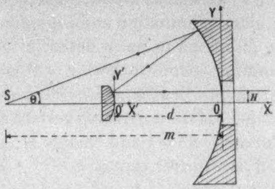


Figure 3.

The objective (Figure 3) consists of a concave primary and a convex secondary mirror. These are separated by a distance d and image parallel light at a distance m from the pole of the primary (m and d are negative in Figure 3). In this case the ray traces are very simple. In the previous notation we have

$$\omega=\Theta, \quad h=-m\tan\Theta; \qquad \omega'=0, \quad h'=H'. \qquad \ldots\ldots(3.1)$$

Corresponding parameters satisfy the sine condition which now takes the form

$$t'=ft, \qquad \ldots\ldots(3.2)$$

where

$$t=\sin\Theta; \qquad t'=H', \qquad \ldots\ldots(3.3)$$

and f is the focal length of the system. On applying the rule for mirrors of §2.3 and making use of the above relations, equations (2.10) and (2.11) take the form

$$\frac{dx}{dt}=-\left[\frac{(R\cos\Theta-R_x)}{(R\sin\Theta-R_y)}+\tan\Theta\right]^{-1}(x-m)\sec^3\Theta, \qquad \ldots\ldots(3.4)$$

$$dx'/dt=fR_y/(R-R_x), \qquad \ldots\ldots(3.5)$$

where as before

$$R_x=x'-x+d, \quad R_y=y'-y, \quad \text{and} \quad R^2=R_x^{\,2}+R_y^{\,2}. \qquad \ldots\ldots(3.6)$$

Also (2.8) and (2.9) become

$$y=(x-m)\tan\Theta \qquad \ldots\ldots(3.7)$$

and

$$y'=f\sin\Theta. \qquad \ldots\ldots(3.8)$$

We computed the system with design data $m=-6$ cm., $d=-3$ cm., $f=0{\cdot}75$ cm. and numerical aperture (N.A.) $=0{\cdot}58$.

Twenty equally spaced values of the parameter t were selected, viz.

$$t_r=0{\cdot}029r \quad (r=1,2,\ldots.20).$$

First the values of x_r and x_r' with the corresponding values of y_r and y_r' were computed for $r=1$ and $r=2$ by Runge's method. The integration was then continued by Adam's method. In calculating the first few differences occurring in Adam's formulae use was made of the fact that $x_{-r}=x_r$ and $x_{-r}=x_r'$. Results for the first six values of x are shown in table 2, where Δ denotes the difference between these solutions and those obtained from Schwarzschild's formulae.

Owing to the inaccuracies of Runge's method an almost constant error of 8×10^{-6} cm. occurred in all values of x_r for $r > 2$. The maximum error in x_r' was 2×10^{-6} cm.

Table 1. Reflecting Microscope, $m = -6$ cm., $d = -3$ cm., $f = 0{\cdot}75$ cm., N.A. = 0 58.

		Concave primary (in cm.)				Convex secondary (in cm.)		
r	$t = \sin \Theta_r$	$-x_r$	$10^6 \Delta x_r$	y_r	$10^6 \Delta y_r$	$-x_r'$	$10^6 \Delta x_r'$	$y_r' = H'$
1	0·029	0·003469	0	0·173973	0	0·000276	0	0·02175
2	0·058	0·013880	0	0·347781	0	0·001104	1	0·04350
3	0·087	0·031240	0	0·521258	1	0·002486	0	0·06525
4	0·116	0·055562	1	0·694241	1	0·004424	0	0·08700
5	0·145	0·086866	1	0·866562	1	0·006920	0	0·10875
6	0·174	0·125175	0	1·038054	0	0·009978	1	0·13050
7	0·203	0·170523	0	1·208547	0	0·013603	1	0·15225
8	0·232	0·222949	1	1·377870	0	0·017799	−1	0·17400
9	0·261	0·282499	1	1·545849	0	0·022575	−1	0·19575
10	0·290	0·349231	0	1·712307	0	0·027937	−1	0·21750
11	0·319	0·423213	0	1·877063	0	0·033894	0	0·23925
12	0·348	0·504524	0	2·039933	0	0·040456	0	0·26100
13	0·377	0·593259	0	2·200726	0	0·047634	0	0·28275
14	0·406	0·689530	0	2·359245	0	0·055441	0	0·30450
15	0·435	0·793468	−1	2·515288	−1	0·063892	1	0·32625
16	0·464	0·905230	0	2·668638	−1	0·073002	2	0·34800
17	0·493	1·024999	−1	2·819072	0	0·082790	1	0·36975
18	0·522	1·152996	0	2·966351	0	0·093277	−1	0·39150
19	0·551	1·289480	−2	3·110222	−1	0·104487	0	0·41325
20	0·580	1·434771	−1	3·250403	1	0·116445	−1	0·43500

Δ denotes difference between results obtained from Schwarzchild's formulae and those calculated by the present methods.

Table 2

r	1	2	3	4	5	6
$-x_r$	0·003471	0·013884	0·031248	0·055570	0·086874	0·125183
$10^6 \Delta x_r$	2	4	8	9	9	8

To obtain higher accuracy we repeated the integration for the first two points. This was carried out by substituting the values of dx_r/dt and dx_r'/dt $(r = 0, 1, 2)$ previously obtained into the formulae of Newton-Cotes (Whittaker and Robinson 1946, p. 152). Very close agreement with Schwarzschild's results was obtained. The integration was then continued as before with results given in Table 1. It can be seen that Δ never exceeds 2×10^{-6} cm., i.e. about 1/20 of a reflection fringe. Such high accuracy in the design data is not needed by the practical optician, but it is a necessary preliminary to a numerical analysis of the off-axis errors, on which any theoretical assessment of the value of the design must be based.

The computation was carried out with the aid of Gifford's 8-figure tables of trigonometric functions (interpolating up to 1/100″) and Milne-Thomson's 8-figure tables of square roots.

Since, in the derivation of our equations, no use was made of any approximations, any inaccuracies in the final solution are due to small computational errors.

ACKNOWLEDGMENT

We are indebted to Drs. C. R. Burch, H. Fröhlich and E. H. Linfoot for helpful criticisms and encouragement.

REFERENCES

Burch, C. R., 1942, *Mon. Not. R. Astr. Soc.*, **102**, 159; 1947, *Proc. Phys. Soc.*, **59**, 41.
Bureau, F., and Swings, P., 1934, *Rev. d'Optique*, **13**, 127.
Chrétien, H., 1922, *Rev. d'Optique*, **1**, 49.
Herzberger, M., and Hoadley, H. O., 1946, *J. Opt. Soc. Amer.*, **36**, 334.
Linfoot, E. H., 1943, *Proc. Phys. Soc.*, **55**, 481; 1944, *Mon. Not. R. Astr. Soc.*, **104**, 48.
Martin, L. C., 1944, *Proc. Phys. Soc.*, **56**, 104.
Runge, C., and Konig, H., 1924, *Numerisches Rechnen* (Berlin: Springer).
Schwarzschild, K., 1905, *Göttingen Observatory*.
Volosov, D. S., 1947, *J. Opt. Soc. Amer.*, **37**, 342.
Whittaker, E. T., and Robinson, G., 1946, *The Calculus of Observations* (London: Blackie & Son).
Wolf, E., 1948, *Proc. Phys. Soc.*, **61**, 494.

The Refractive Index in Electron Optics and the Principles of Dynamics

By W. EHRENBERG * and R. E. SIDAY †

* Birkbeck College, University of London † I.C.I. Fellow, Edinburgh University

MS. received 23rd February 1948, *and in amended form 28th July* 1948; *read 3rd December* 1948

ABSTRACT. In view of mis-statements made in the literature, the origin of the refractive index in electron optics is discussed in some detail, and the uniqueness of an expression previously given is demonstrated. On this basis, some general properties of electron optics are investigated.

A relation between ray direction and wave normal is obtained. Whereas the refractive index is unique in terms of the magnetic vector potential **A**, this itself is arbitrary to some extent. It is shown that **A** must, for purposes of electron optics, be chosen so as to satisfy Stokes' theorem and that, if it does, no observable effects result from the arbitrariness of **A**. An expression for the optical path difference is given in terms of the magnetic flux enclosed. The results are applied to a number of questions, viz. the differential equations for trajectories, the focusing properties of an axially symmetric field and the interference pattern produced by two converging bundles of rays which enclose a magnetic flux.

§ 1. INTRODUCTION

Whereas in light optics the refractive index of a medium is in the first instance an experimental datum, and its accurate value is the basis of any detailed discussion of the performance of optical instruments, all geometrical electron optics is entirely contained in Lorentz's equation for the forces acting on a moving charge. As a result, all equations for trajectories can be derived directly from that equation by specifying the electric and magnetic fields. Thus the rôle of the refractive index in electron optics is far less obvious than that of its counterpart for light, and in fact different authors have proposed essentially different values of the refractive index for the same field without arousing much perturbation (Glaser 1933, 1937, Opatowski 1943). Only through the persistent use of the refractive index, however, is the optical character of electron optics fully brought out, the flow of reasoning maintained and in fact greatly simplified, and the way opened for a wave optical treatment of electronic problems, such as the resolving power of the electron microscope.

It is Glaser's great merit to have first pointed out the existence of this refractive index and to have given its value correctly. The reasoning that has led him to it is, however, as will be shown, inexact, so that a logical extension of his line of thought has led to wrong values (Opatowski 1943). It is the object of this paper to review the source of the refractive index in electron optics and to discuss some of its general properties which in view of its peculiar indefiniteness require examination.

§ 2. THE DERIVATION OF THE REFRACTIVE INDEX

The term "electron optics" is often justified by reference to the correspondence between instruments designed for electrons and for light. A systematic definition of optics in a generalized sense appears to arise from "Fermat's principle". Although in its historical form it is a principle of fastest arrival, it is now recognized as

$$\overset{P_1}{\underset{P_0}{\delta}}\int \mu\, ds = 0. \qquad \ldots\ldots(1)$$

Here ds stands for the length of an element of ray, μ denotes the refractive index, in general a function of position, direction and frequency, $\int \mu\, ds$ is the line integral over any line between the limits indicated as subscripts and superscripts of δ.*

Fermat's principle may be enunciated as follows: In any optical medium a scalar quantity μ, finite everywhere in space, can be defined so that the line integral in the three-dimensional space taken between any fixed points is an extremum for the ray which passes through these points. Or, more precisely: If the value of $\int_{P_0}^{P_1} \mu\, ds$ along a particular line joining any two points P_0 and P_1 of an optical system is stationary with respect to weak variations that leave the end points unchanged, then the line is a ray of the optical system. As an obvious corollary it follows that: If no line can be drawn joining P_0 and P_1, such that $\int \mu\, ds$ taken along this line is stationary for weak variations, no ray can pass according to geometrical optics from P_0 to P_1.

In the optics of light the value of μ as a measured quantity is finite and single valued and is continuous except at a finite number of surfaces separating different media. As will become apparent later, however, more general mathematical expressions can be used in place of the simple values usually given, and in electron optics they arise in fact automatically. Particular expressions are admissible only if they do not violate the conditions implied for the validity of Fermat's result. Firstly, the difference of Fermat's integrals along any two paths must be defined anywhere in space where the optical properties are to be investigated. This requires that the refractive index should be fixed everywhere in space once it is fixed in the neighbourhood of one point, so that μ must be single valued. Secondly, it should have no singularities through which Fermat's integral would become infinite. And thirdly, any discontinuities should be of such a nature that they appear as limiting cases of a continuous μ. This is true for the boundary between two optical media but would not be true if, for example, μ were proportional to a cyclic coordinate which is limited to the values 0 to 2π in order to make it single-valued, for the numerical values of the integrals along either side of an irremovable discontinuity will not be equal even when the lines are infinitely close together. That would make the weak variation required by Fermat's principle impossible.

---

* Sometimes the variational operator is written δ for a variation in which the limits of the variable of integration are to be kept constant, and Δ if these limits should be varied. Sometimes the boundary conditions are not given in the equation but must be found from the context. We find this confusing and propose to write the complete boundary conditions at the lower and upper limit of the integral below and above the δ sign, so that, for instance, x_0, y_0, z_0, t_0 below δ denote these values for the line at the lower limit etc. $x_0, y_0, z_0, t_0+\Delta t_0$ denote that at the lower limit $x=x_0$, $y=y_0$ etc. but that the value for t_0 should be varied; or in other words, that the lines under comparison must go through x_0, y_0, z_0, but may do so at any value t near t_0. We shall also use the symbol P_0 to denote the set of coordinates x_0, y_0, z_0 and P_1 for x_1, y_1, z_1.

Fermat's principle does not contain time; and, in fact, not only in geometrical optics (Gabor 1945) but also in wave optics the timing of the phenomena observed is of no interest. Problems into which timing enters, such as the velocity of light, stand quite apart. Thus Fermat's principle suggests a generalized optics whenever problems arise in any field in which timing is of no interest and a function of position and direction can be defined so that rays are determined by an expression of the form of Fermat's principle.

It is now important to see that such a generalization cannot be based on Hamilton's principle

$$\delta \int_{P_0,t_0}^{P_1,t_1} L\,dt = 0 \qquad \ldots\ldots(2)$$

in spite of the fact that its integrand is readily transformed into the form of Fermat's integrand. Hamilton's principle may be enunciated: The world line of a particle (i.e. the line in the x, y, z, t space) which goes through the points P_0, t_0 and P_1, t_1 is the line going through these points which minimizes the line integral $\int_{t_0}^{t_1} L\,dt$. Its general validity arises from its equivalence with the equations of motion of Lagrange:

$$\frac{\partial L}{\partial q_i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) = 0. \qquad \ldots\ldots(3)$$

For electrons in particular,

$$L = \frac{mv^2}{2} - \mathbf{e}\Phi + \frac{\mathbf{e}}{\mathbf{c}}(\mathbf{vA}) \qquad \ldots\ldots(4)$$

(Schwarzschild 1903). It is only necessary to write this L in, say, Cartesian coordinates x, y, z, $\dot{x}$, $\dot{y}$, $\dot{z}$, to perform the differentiations (3) and to insert $-\operatorname{grad}\Phi = \mathbf{E}$, $\operatorname{curl}\mathbf{A} = \mathbf{H}$ in order to arrive at Lorentz' equation for the acceleration of an electron

$$m\dot{\mathbf{v}} = \mathbf{e}\mathbf{E} + \frac{\mathbf{e}}{\mathbf{c}}[\mathbf{v} \times \mathbf{H}] \qquad \ldots\ldots(5)$$

in Gauss' units, with $\mathbf{e} = -4{\cdot}8 \times 10^{-10}$ E.S.U.

That Fermat's principle cannot be associated with Hamilton's principle was, of course, clearly realized by L. de Broglie, but later authors who have specialized in electron optics have tried to associate the two principles by equating $\mu = L/v$. Only for a particular—valid but unjustified—choice of L does this lead to a correct answer. Glaser, in fact, by this artifice obtained a correct value, whereas Opatowski, by generalizing Glaser's choice of L, was led to a set of values for μ some of which are wrong. Glaser's correct result appears to be due to such a combination of an arbitrary breaking up of his integrand into two terms with a neglect of

Hamilton's boundary conditions that Lagrange's form of the principle of least action (see below) is in fact substituted for Hamilton's principle. In some textbooks on electron optics (e.g. Maloff and Epstein 1938) the correct variational principle is used while in others it is not. We have found no explicit discussion of this point in any, and the choice of a principle seems to have been based rather on "luck than judgment".

It is quite correct to transform the integrand *Ldt* since a given line $x=x(t)$, $y=y(t)$, $z=z(t)$ which satisfies $x_0=x(t_0)$, $y_0=y(t_0)$, $\dots\ z_1=z(t_1)$ and which gives to the integral $\int_{t_0}^{t_1} L\,dt$ the value B, will give the same value to the integrals $\int_{P_0}^{P_1} (L/v)\,ds$ or $\int_{x_0}^{x_1} (L/\dot{x})\,dx$ etc. so that Hamilton's principle can alternatively be written

$$\underset{P_0,t_0}{\overset{P_1,t_1}{\delta}} \int (L/v)\,ds=0, \qquad \dots\dots(6)$$

$$\underset{t_0,y_0,z_0,x_0}{\overset{t_1,y_1,z_1,x_1}{\delta}} \int (L/\dot{x})\,dx=0 \text{ etc.}, \qquad \dots\dots(6\,a)$$

where in (6 *a*) $L/\dot{x}$ is now a function of x, y, z, dy/dx, dz/dx, dt/dx. Further, it is easily seen that equations (3) can be combined to

$$(d/dt)\,(L-\Sigma\dot{q}_i(\partial L/\partial\dot{q}_i))-\partial L/\partial t=0 \qquad \dots\dots(7)$$

so that

$$L-\Sigma\dot{q}_i\partial L/\partial\dot{q}_i=\text{const.}=h \qquad \dots\dots(7\,a)$$

($-h=$ the energy of the electron) if L does not depend explicitly on t, i.e. if the magnetic and the electric fields are constant in time—the case we are here exclusively interested in. As

$$\dot{x}=v(dx/ds)=v/\sqrt{(1+y'^2+z'^2)}, \quad \dot{y}=y'\dot{x}, \quad \dot{z}=z'\dot{x}, \quad y'=dy/dx \text{ etc.}, \qquad \dots\dots(8)$$

we can eliminate the variables $\dot{x}$, $\dot{y}$, $\dot{z}$, v from L/v by means of (7 *a*) and (8) and replace them by y', z', h so that the integral in (6) takes the form

$$\int f(x,y,z,y',z',h)\,ds. \qquad \dots\dots(9)$$

This does not contain t or derivatives of the form d/dt and is, in fact, a line integral in three dimensional space; that is, it depends only on the projection of the world line on three dimensional space, with h as a parameter, as the analogy to Fermat's principle would require.

But the value of h in this integrand can be determined numerically only after Lagrange's equations have actually been solved and the boundary conditions inserted. For equation (7 *a*) is not an equation for h; it restricts the lines under test by (6), as an equation between x, y, z, $\dot{x}$, $\dot{y}$, $\dot{z}$, h, in a way analogous to the restriction that would be imposed by an equation $x=x(t)$. Only for one particular value of h is the world line which minimizes (2) or (6) contained in a set restricted by (7 *a*). Moreover, equations (2) or (6) require the integral to be an extremum *for given limits in space and time.* If we now minimize (9), we obtain a minimum *regardless of time limits,* as the value of the integral does not depend on time any longer. The minimum of (9) will not in general (i.e. for arbitrary h) pass the end points in space at the times t_0 and t_1. Thus, although now the integral in (6) has the form $\int \mu\,ds$, its minimum is not that required by Hamilton's principle.

It is thus seen that the analogy between Fermat's and Hamilton's principles fails, because the lines under comparison in Hamilton's principle are not lines of constant energy which can be fixed arbitrarily, and because the extremum required is not an extremum regardless of the duration of transit between local points, such as would be found by considering the projection of the world line on space only.

As pointed out by L. de Broglie (1930) the problem of generalized optics is in fact solved by the principle of least action that has developed from Maupertuis' principle. Following Whittaker (1917) * it may be enunciated as follows: The equation

$$\delta \int_{P_0,\, t_0+\Delta t_0}^{P_1,\, t_1+\Delta t_1} \left(\Sigma \dot{q}_i \frac{\partial L}{\partial \dot{q}_i}\right) dt = 0 \qquad \ldots\ldots(10)$$

leads to the equations of Lagrange if the lines under examination are subjected to the condition

$$L - \Sigma \dot{q}_i (\partial L / \partial \dot{q}_i) = h; \quad h \text{ arbitrary.} \qquad \ldots\ldots(11)$$

If now (10) is written

$$\delta \int_{P_0,\, t_0+\Delta t_0}^{P_1,\, t_1+\Delta t_1} \frac{1}{v} \left(\Sigma \dot{q}_i \frac{\partial L}{\partial \dot{q}_i}\right) ds = 0, \qquad \ldots\ldots(12)$$

one can replace $\dot{x}$, $\dot{y}$, $\dot{z}$, v in the integrand by h and purely spatial derivatives as before, allowing at the same time for the condition (11). The line integral is then an integral in space only and the open boundary conditions for (12) are satisfied by disregarding the t limits. Hence the principle becomes

$$\delta \int_{P_0}^{P_1} \mu \, ds \qquad \ldots\ldots(13)$$

with

$$\mu = (1/v) \Sigma \dot{q}_i (\partial L / \partial \dot{q}_i). \qquad \ldots\ldots(14)$$

With L given by (4),

$$\mu = \gamma \{ \boldsymbol{m} v + (\boldsymbol{e}/\mathbf{c})(\mathbf{SA}) \}, \qquad \ldots\ldots(15)$$

where γ is an arbitrary constant, $\mathbf{S}$ is the unit vector in the direction of the line, v is the velocity of the electron which, as a function of position only, is given by (11), viz.

$$-\tfrac{1}{2} \boldsymbol{m} v^2 - \boldsymbol{e}\Phi = h, \quad \text{i.e. } v = \sqrt{\{-(h + \boldsymbol{e}\Phi)(2/\boldsymbol{m})\}} \qquad \ldots\ldots(16)$$

and h is a constant which can be determined, e.g. by requiring v to have a given value at a point of given Φ. Then

$$\mu = \gamma [\sqrt{\{-2\boldsymbol{m}(h + \boldsymbol{e}\Phi)\}} + (\boldsymbol{e}/\mathbf{c})(\mathbf{SA})]. \qquad \ldots\ldots(15\,a)$$

If the electric potential is constant, i.e. in the purely magnetic case, v is constant and need not be replaced by (16). Alternatively we can express v in terms of $(H\rho)$, viz. $\boldsymbol{m} v = (\boldsymbol{e}/\mathbf{c}) H\rho$, when

$$\mu = \gamma (\boldsymbol{e}/\mathbf{c}) \{H\rho\} \{1 + (\mathbf{SA})/(H\rho)\} \qquad \ldots\ldots(15\,b)$$

which in the purely magnetic case is equivalent to

$$\mu = 1 + (\mathbf{SA})/(H\rho) \qquad \ldots\ldots(15\,c)$$

---

* The proof given by Whittaker is especially adapted to mechanics. There are a number of less general formulations of the principle which do not lend themselves readily to the present problem. A proof of the equivalence of the principle of least action to Lagrange's equations, based on general methods of the variational calculus, is given by Bolza (1909).

(15 *b*) is also relativistically correct. Using the notation $\partial L/\partial \dot{q}_i = p_i$ where the p_i are the components of the generalized momentum, the integral in (10) may be written

$$\int (\mathbf{p}\dot{\mathbf{q}})dt = \int (\mathbf{pS})ds. \qquad \ldots\ldots(17)$$

Hence it is seen that in general $\mu = (\mathbf{pS})$ where $\mathbf{p}$ is expressed in terms of the coordinates, direction cosines and spatial derivatives by means of (11).

§ 3. THE TRANSITION TO WAVE OPTICS: RELATION BETWEEN THE RAY AND THE WAVE NORMAL

Although, as explained in the introduction, the geometrical optics is completely described by the differential equations of motion, for the development of the physical optics of a system it is necessary to regard the trajectories as rays defined by the refractive index and Fermat's principle. For then the well-known surfaces of Malus associated with families of rays can be established and these in turn are to be identified with the wave surfaces of physical optics. This is the transition to wave optics. The presence of the term (**SA**) in the expressions (15) for the refractive index in electron optics shows that the electron optical properties of magnetic fields are those of an optically inhomogeneous and anisotropic medium. The surfaces of Malus will therefore not be normal to the rays but they must still be identified with the wave surfaces.

In an instructive paper, Frank (1933) has shown how the differential equations of a ray in such cases can be expressed by a vector equation (thereby extending the method and results of Sommerfeld and Runge (1911)), and has shown how the relation between the ray and the wave normal can be determined. The latter relation will now be established here directly from Fermat's integral by extending the analysis frequently given for isotropic media (e.g. Forsyth 1927, pp. 229, 256–262) to the case of anisotropic media.

In order to link up geometrical optics with physical optics it is necessary to consider families of rays, viz. the totality of rays emerging from one point. For such a family

$$W = \int_{P_0}^{P_1} \mu\, dr, \qquad \ldots\ldots(18)$$

where *dr* is an element of path along a ray, has a definite value at any point P_1 in space through which a ray coming from P_0 passes. *W* expressed as a function of this point (i.e. the end point) is called Hamilton's characteristic function, and the surfaces $W = \text{constant}$ are called the surfaces of Malus.

Let us commence by considering a family of non-intersecting lines (in general not rays) all emerging from P_0 and filling the space continuously. They may be represented by a family of equations

$$x = x_k(u), \qquad y = y_k(u), \qquad z = z_k(u), \qquad \ldots\ldots(19)$$

where *k* is a parameter and for all lines $x(0) = x_0$ etc. Then each point P_1 in space is defined by a particular set of equations, and a particular value of the auxiliary variable *u*. And, conversely,

$$W' = \int_0^{u'} g\, du, \qquad \ldots\ldots(20)$$

has a definite value at every point in space, if *g* is defined by

$$g = \mu\lambda, \qquad \ldots\ldots(21)$$

where the refractive index μ, originally a function of x, y, z and the not independent direction cosines dx/ds, dy/ds, dz/ds, has become in (20) a function of the independent differential coefficients

$$x_1 = dx/du, \qquad y_1 = dy/du, \qquad z_1 = dz/du, \qquad \ldots\ldots(22)$$

viz. $\mu = \mu(x, y, z, x_1/\lambda, y_1/\lambda, z_1/\lambda)$, and λ is given by

$$\lambda = ds/du = \sqrt{(x_1{}^2 + y_1{}^2 + z_1{}^2)}. \qquad \ldots\ldots(23)$$

Because of its origin the function g satisfies the permanent identity

$$g = x_1(\partial g/\partial x_1) + y_1(\partial g/\partial y_1) + z_1(\partial g/\partial z_1), \qquad \ldots\ldots(24)$$

so that $dW' = gdu$ from (20) becomes

$$dW' = x_1 \frac{\partial g}{\partial x_1} du + y_1 \frac{\partial g}{\partial y_1} du + z_1 \frac{\partial g}{\partial z_1} du = \frac{\partial g}{\partial x_1} dx + \frac{\partial g}{\partial y_1} dy + \frac{\partial g}{\partial z_1} dz. \qquad \ldots\ldots(25)$$

Hence

$$\partial W'/\partial x = \partial g/\partial x_1 = n'_x; \quad \partial W'/\partial y = \partial g/\partial y_1 = n'_y; \quad \partial W'/\partial z = \partial g/\partial z_1 = n'_z, \qquad \ldots\ldots(26)$$

or $\mathbf{n}' = \text{grad } W'$, i.e. the vector $\mathbf{n}'$ defined by its components lies perpendicular to the system of surfaces $W' = \text{constant}$.

Now from (21)

$$dg = \mu d\lambda + \lambda d\mu = (\mu/\lambda)(x_1 dx_1 + y_1 dy_1 + z_1 dz_1)$$
$$+ \lambda \left(\frac{\partial \mu}{\partial x} dx + \frac{\partial \mu}{\partial y} dy + \frac{\partial \mu}{\partial z} dz + \frac{\partial \mu}{\partial(x_1/\lambda)} d\left(\frac{x_1}{\lambda}\right) + \frac{\partial \mu}{\partial(y_1/\lambda)} d\left(\frac{y_1}{\lambda}\right) + \frac{\partial \mu}{\partial(z_1/\lambda)} d\left(\frac{z_1}{\lambda}\right) \right) \qquad \ldots\ldots(27)$$

and

$$d(x_1/\lambda) = dx_1/\lambda - (x_1/\lambda)(x_1 dx_1 + y_1 dy_1 + z_1 dz_1) \text{ etc.} \qquad \ldots\ldots(28)$$

Thus

$$n'_x = \frac{\partial g}{\partial x_1} = \mu \frac{x_1}{\lambda} - \lambda \frac{\partial \mu}{\partial(x_1/\lambda)} \left(\frac{1}{\lambda} - \frac{x_1{}^2}{\lambda^3} \right) - \lambda \frac{\partial \mu}{\partial(y_1/\lambda)} \frac{x_1 y_1}{\lambda^3} - \lambda \frac{\partial \mu}{\partial(z_1/\lambda)} \frac{x_1 z_1}{\lambda^3}$$
$$= \mu S_x - \frac{\partial \mu}{\partial S_x} - \frac{S_x}{\lambda} \left(x_1 \frac{\partial \mu}{\partial S_x} + y_1 \frac{\partial \mu}{\partial S_y} + z_1 \frac{\partial \mu}{\partial S_z} \right). \qquad \ldots\ldots(29)$$

In the last expression S_x, S_y, S_z denote the components dx/dS etc. of the unit vector along the lines which by (22) and (23) have the values x_1/λ, y_1/λ, z_1/λ.

Up to this point the lines $x = x(u)$ etc. have been arbitrary, and in consequence also the surfaces $W' = \text{constant}$. Now we specialize: We let the lines be rays so that the unit vectors (S_x, S_y, S_z) along the lines become the unit vectors (R_x, R_y, R_z) along the rays and we can identify the auxiliary variable u with the length of the rays measured from the origin so that $\lambda = 1$. Then $\mathbf{n}$ which we write now for $\mathbf{n}'$ is given by its components

$$n_x = \frac{\partial g}{\partial R_x} = \mu R_x + \frac{\partial \mu}{\partial R_x} - R_x \left(R_x \frac{\partial \mu}{\partial R_x} + R_y \frac{\partial \mu}{\partial R_y} + R_z \frac{\partial \mu}{\partial R_z} \right) \qquad \ldots\ldots(30)$$

and W' becomes W now given by

$$W = \int_{P_0}^{P_1} \mu \, dr, \qquad \ldots\ldots(18)$$

the integrals taken along the *rays*. That means, the surfaces $W=\text{constant}$ have become the surfaces of Malus. Equations (26) show that the vector $\mathbf{n}$ defined by (30) is normal to the surfaces $W=\text{constant}$ so that (30) gives the general relation between the surface normals and the direction of the ray. They are inclined at an angle ξ given by $\cos\xi=(\mathbf{R},\ \mathbf{n}/n)$. By (15), μ in the direction of the ray is given by

$$\gamma(\boldsymbol{m}v+(\mathbf{e}/\mathbf{c})(R_xA_x+R_yA_y+R_zA_z),$$

so that from (30)

$$\mathbf{n}=\gamma(\boldsymbol{m}v\mathbf{R}+(\mathbf{e}/\mathbf{c})\mathbf{A}), \qquad \ldots\ldots(31)$$

a result given by Glaser (1933). It is seen that $\mathbf{n}$ is proportional to the generalized momentum of the electron.* Hence

$$\frac{\mathbf{n}}{n}=\frac{\boldsymbol{m}v\mathbf{R}+(\mathbf{e}/\mathbf{c})\mathbf{A}}{\sqrt{\boldsymbol{m}^2v^2+(\mathbf{e}/\mathbf{c})A^2+2\boldsymbol{m}v(\mathbf{e}/\mathbf{c})A\cos\theta}}, \qquad \ldots\ldots(32)$$

where θ is the angle between $\mathbf{A}$ and the ray. Then

$$\cos\xi=\frac{\boldsymbol{m}v+(\mathbf{e}/\mathbf{c})A\cos\theta}{\sqrt{\boldsymbol{m}^2v^2+(\mathbf{e}/\mathbf{c})^2A^2+2\boldsymbol{m}v(\mathbf{e}/\mathbf{c})A\cos\theta}}=\sqrt{1+\left(\frac{(\mathbf{e}/\mathbf{c})A\sin\theta}{\boldsymbol{m}v+(\mathbf{e}/\mathbf{c})A\cos\theta}\right)^2}$$

giving $\qquad \ldots\ldots(33)$

$$\xi=\tan^{-1}\left(\frac{(\mathbf{e}/\mathbf{c})A\sin\theta}{\boldsymbol{m}v+(\mathbf{e}/\mathbf{c})A\cos\theta}\right). \qquad \ldots\ldots(34)$$

Geometrical optics is expressed through the value of the refractive index and Fermat's principle alone. The latter assumes that rays can be defined. Wave optics is expressed through the refractive index, Huyghens' principle and the wavelength. In general the descriptions are different, in which case that given by geometrical optics is wrong. Under certain circumstances, however, they lead to the same result and when, but only when, this is true, light can be said to travel along the rays of geometrical optics, and Fermat's principle can be deduced from that of Huyghens. Then the surfaces of Malus can be identified with the wave surfaces of physical optics and (34) gives the angle between the wave normal and the direction of the ray. At the same time the refractive index of geometrical optics becomes inversely proportional to the wavelengths.

This description is based purely on classical physics. To complete the description of any electron-optical phenomenon it is necessary to introduce the absolute size of the wavelength. In the case of an isotropic medium, in the absence of a magnetic vector potential, the well-known formula of L. de Broglie

$$\lambda_0=\boldsymbol{h}/\boldsymbol{m}v=\boldsymbol{ch}/\boldsymbol{e}(H\rho) \qquad \ldots\ldots(35)$$

is valid.

As in general

$$\lambda/\lambda_0=\mu_0/\mu, \qquad \ldots\ldots(36)$$

equation (15 *b*) yields at once

$$\lambda=\lambda_0/\{1+(\mathbf{SA})/(H\rho)\}. \qquad \ldots\ldots(37)$$

Here λ is the actual wavelength, expressed in terms of the wavelength λ_0 which the electron would have at the same point in space in the absence of $\mathbf{A}$, or at a point in space of the same electrostatic potential where $\mathbf{A}=0$.

---

* So that curl $\mathbf{p}$ always vanishes for a family of rays emerging from a point. Gabor (1945) has shown that curl $\mathbf{p}$ also vanishes for a family of rays issuing from an extended source, under certain restrictions, and hence, that " surfaces of Malus " can be constructed even in this case which has no analogy in light optics. This phenomenon is associated with the existence of the first cross-over in front of thermionic cathodes.

§4. THE OPTICAL SIGNIFICANCE OF THE ARBITRARINESS OF THE MAGNETIC VECTOR POTENTIAL

The detailed discussion of the origin of μ given above shows clearly the extent to which the mapping of μ throughout space is arbitrary. The expressions (15) as written down are unique. An arbitrary additive constant in L would have dropped out. Therefore the only freedom left in assigning a numerical value to μ arises from an arbitrary additive constant F included in Φ and the gradient of arbitrary scalar ψ included in $\mathbf{A}$: $\mathbf{A}=\mathbf{A}_0+\operatorname{grad}\psi$. Hence in general

$$\mu=\gamma\{[-2\boldsymbol{m}(h+\boldsymbol{e}\Phi_0+F)]^{\frac{1}{2}}+(\boldsymbol{e}/\mathbf{c})(\mathbf{SA})+(\boldsymbol{e}/\mathbf{c})(\mathbf{S}\operatorname{grad}\psi)\}, \quad \ldots\ldots(15\,d)$$

where the first term denotes the isotropic contribution to μ arising from the electrostatic potential and the second and third terms the anisotropic contributions. F combines with the arbitrary constant of integration and produces thus no new effects at all when h is determined by the boundary conditions. There is no general term k/v (k an arbitrary constant) as proposed by Opatowski in 1943. The term containing $\operatorname{grad}\psi$ to which he has rightly drawn attention is, however, of considerable interest.

Electrodynamic considerations require only that $\operatorname{curl}\mathbf{A}$ reproduces a given field $\mathbf{H}$ everywhere. But $(\mathbf{SA})$ occurs as an additive term in the refractive index and it was seen in §2 that any expression for the refractive index is subject to the condition that it is continuous and free from singularities. Hence $(\mathbf{SA})$ and therefore $\mathbf{A}$ must satisfy the same condition if it is to play its rôle in electron optics. These conditions are just those for the validity of Stokes' theorem (Kellogg 1929, p. 91). The further condition of the continuity of the first derivative need not be specially considered in physics as it can always be satisfied by a physically insignificant alteration in the vector potential.

Hence it appears that the vector potential in electron optics must satisfy Stokes' theorem, and this is the only valid restriction which must be imposed on it. In consequence $\mathbf{A}$ cannot in general be chosen so as to vanish with the magnetic field. An example will be given in the last section.

In the electromagnetic theory it is customary so to choose $\mathbf{A}$ that for static fields $\operatorname{div}\mathbf{A}=0$.

This condition is frequently applied to the vector potential in electron optics, as, for example, in the book by Zworykin *et al.* (1945). It is readily seen, however, from the development given, that in this connection such a condition is entirely arbitrary. Further, it is not useful as it does not fix the value of $\mathbf{A}$ uniquely, for the value of ψ can still be chosen to be any solution of Laplace's equation $\operatorname{div}\operatorname{grad}\psi=0$, and is therefore not uniquely defined. The remaining arbitrariness in $\mathbf{A}$, or $\operatorname{grad}\psi$ if an $\mathbf{A}_0$ satisfying Stokes' theorem is given, is still considerable.

As the arbitrariness of the vector potential does not enter Lorentz' equation it cannot produce any observable effects in geometrical optics. This can, of course, also be shown by a consideration of Fermat's principle with the value for μ given. But equation (34) shows that the inclination ξ of the wave surfaces to the rays depends on the absolute value A of the vector potential and on its inclination θ to the rays. As the addition of a term $\operatorname{grad}\psi$ to $\mathbf{A}_0$ leaves the rays unchanged but changes $\mathbf{A}$ with respect to both direction and absolute value, both A and θ depend on the arbitrary ψ. The form of (34) shows that in general the change of A is not compensated by the change of θ, so that in general the addition of a term $\operatorname{grad}\psi$ to $\mathbf{A}_0$ results in a change of ξ, and thus, as the rays remain unaltered, in a change of the wave surfaces. Hence by a proper choice of ψ an anisotropic medium can be made isotropic locally, and an isotropic medium can be made anisotropic, in an infinite variety of ways.

It is seen, therefore, that the extra term $(\mathbf{e}/\mathbf{c})(\mathbf{S}\,\mathrm{grad}\,\psi)$ profoundly alters the characteristics of the electron waves, but it will now be shown that it does so in such a way that there are no consequences of this change that could be observed in any experiment. The basic problem of physical optics is the evaluation of the optical path lengths of pairs of interfering rays connecting given points. Apart from questions of intensity, the physical phenomena observed depend only on the optical path difference for such pairs. Now the optical path length along a given line connecting a point 0 and 1 is given by

$$\int_0^1 \mu\,ds = \int_0^1 [\boldsymbol{mv} + (\mathbf{SA})]\,ds, \qquad \ldots\ldots(38)$$

for a particular choice of $\mathbf{A}$, viz. $\mathbf{A}_0$. For another value of $\mathbf{A}$ given by $A_0 + \mathrm{grad}\,\psi$ the optical path lengths will be different from (38) by $\int_0^1 (\mathbf{S}\,\mathrm{grad}\,\psi)\,ds = \psi_1 - \psi_0$. As ψ is single valued this difference is the same for each ray connecting the two points considered. Thus the physical result of any electron-optical calculation and in particular any interference effects is independent of how $\mathbf{A}$ is chosen but the detailed picture of the relation between rays and wave surfaces is not. The particular set of wave surfaces belonging to a family of rays and the inclination of the wave surfaces to the ray at that point will vary according to the choice of $\mathbf{A}$. This is compensated, however, by the change in the wavelength given by equation (37). By inserting reasonable numerical values for $\mathbf{A}$ and $(H\rho)$ it is readily seen that the factor of λ_0 may be considerably different from unity, so that the wavelength of an electron beam is largely arbitrary.

Since the physical and geometrical character of a solution of an electron-optical problem is independent of the particular choice of $\mathbf{A}$, one may expect to find a solution of any electron-optical problem in terms of uniquely defined quantities. Such a quantity is the magnetic field $\mathbf{H}$, and it is easy to see how problems can be solved in terms of $\mathbf{H}$ rather than $\mathbf{A}$, for the solution requires only that it should be possible to evaluate the difference d in the values of $\int \mu ds$ along any two paths between the object point and the chosen point of the image space. With μ given by (15 *b*)

$$d = (\mathbf{e}/\mathbf{c})\gamma\left[\int_0^1 (H\rho)\,ds - \int_0^1 (H\rho)\,ds' + \int_0^1 (\mathbf{SA})\,ds - \int_0^1 (\mathbf{SA})\,ds'\right], \qquad \ldots\ldots(39)$$

where ds denotes elements of the first and ds' those of the second path. Now,

$$\int_0^1 (\mathbf{SA})\,ds - \int_0^1 (\mathbf{S}'\mathbf{A})\,ds' = \oint (\mathbf{A}\,d\mathbf{S}) = \iint H_n\,d\sigma, \qquad \ldots\ldots(40)$$

where the last integral but one is taken round the closed path formed by the two lines under consideration and the last one arises from it by Stokes' theorem and is the integral of the normal component of $\mathbf{H}$ over any surface bounded by the two lines. Hence

$$d = (\mathbf{e}/\mathbf{c})\gamma\left(\int_0^1 (H\rho)\,ds - \int_0^1 (H\rho)\,ds' + \iint H_n\,d\sigma\right). \qquad \ldots\ldots(41)$$

In general $(H\rho)$ is a function of the coordinates; in the absence of an electric field $(H\rho)=\text{const.}$, when

$$d=(\mathbf{e}/\mathbf{c})\gamma(H\rho)\left(l_0-l_1-\frac{1}{(H\rho)}\iint H_n d\sigma\right) \qquad \ldots\ldots(42)$$

and the order of interference for the two lines comes to

$$\frac{1}{\lambda_0}\left(l_0-l_1+\frac{1}{(H\rho)}\iint H_n d\sigma\right), \qquad \ldots\ldots(43)$$

where l_0 and l_1 are the length of the two lines joining P_0 and P_1.

An elementary illustration of (43) is the deflection of a parallel bundle of electrons by a thin magnetic field of strength $\mathbf{H}$ and thickness b. If the bundle is deflected by a small angle ϕ, two rays d cm. apart acquire a path difference ϕd which by (43) must be compensated by $\{1/(H\rho)\}Hbd=Hbd\mathbf{e}/\boldsymbol{m}\mathbf{c}v$ as the deflected beam is still of zero order of interference. Hence $\phi=Hb\mathbf{e}/\boldsymbol{m}\mathbf{c}v$ which is a well-known formula.

§ 5. SPECIAL PROBLEMS

It remains now to show how the refractive index given leads in fact to correct solutions of electronic problems, and to a consistent description of electron-optical phenomena.

A. Differential equations for trajectories: For the solution of particular problems it is convenient to mark out one of the three coordinates q_1, q_2, q_3, e.g. x if the q_i denote cylindrical coordinates x, r, θ. Then generally $ds=\Lambda dq_1$, e.g. for the case mentioned $ds=\sqrt{\{1+(dr/dx)^2+r^2(d\theta/dx)^2\}}dx$. Then Fermat's principle

$$\delta\int_{q_1^0 q_2^0 q_3^0}^{q_1^1 q_2^1 q_3^1}\mu\Lambda dq_1=0 \qquad \ldots\ldots(44)$$

leads at once to two independent differential equations for trajectories

$$\frac{d}{dq_1}\frac{\partial(\mu\Lambda)}{\partial(dq_2/dq_1)}-\frac{\partial(\mu\Lambda)}{\partial q_2}=0; \qquad \frac{d}{dq_1}\frac{\partial(\mu\Lambda)}{\partial(dq_3/dq_1)}-\frac{\partial(\mu\Lambda)}{\partial q_3}=0. \qquad \ldots\ldots(45)$$

The expression for $\mu\Lambda$ is greatly simplified if the field has symmetry, such as axial symmetry. For this particular case $\mu\Lambda=\boldsymbol{m}v\Lambda+(\mathbf{e}/\mathbf{c})A_\theta r(d\theta/dx)$ and the equations (45) lead through a short calculation to the fundamental equations of electron lens optics (see, for example, Zworykin *et al.* 1945, p. 503).

B. It may be of interest to illustrate the discussion of the wave surfaces in electron optics given by reference to a simple two-dimensional example.

Korsunsky (1945) pointed out that if a parallel bundle of rays in a plane were incident on a uniform magnetic field perpendicular to the plane, having a circular boundary so that $H=H_0$ for $0\leqslant r\leqslant a$ and $H=0$ for $r>a$, then these rays will be focused. The field acts as a focusing prism. If the " momentum " $\mathbf{c}\boldsymbol{m}v/\mathbf{e}=(H\rho)$ is such that $(H\rho)=H_0 a$ the prism gives a perfect focus over its entire aperture; this focusing action is illustrated in Figure 1. The rays are straight lines outside the field and circles of radius a inside it. In this case the optical paths of all the rays arriving at the point I should be equal, and that they are is easily tested using equation (42).

Any ray A_1BI of radius (in the magnetic field) a, can be distinguished by the parametric angle ξ, and any point P on the ray by either the angle θ or the angle ϕ constructed as shown in the figure. θ increases outside the field from 0 to ξ corresponding to the point B where the electron enters the magnetic field, and inside the field from ξ to $\pi/2$ at I, while ϕ increases inside the field from zero to

$2\phi_c = \pi/2 - \xi$ at I. The angle ϕ_c corresponds to the closest approach of P to the axis O of the magnetic field.

Then, using equation (42) but leaving out the constant factor $(\mathbf{e}/\mathbf{c})\gamma(H\rho)$, the optical paths A_1BPI and A_0I *coming from infinity* differ by the difference of their actual lengths plus $H_0\sigma/(H\rho) = \sigma/a$ where σ is the area between the arc BPI and the edge of the circle. The actual lengths are $l_\infty + a(1-\cos\xi) + 2a\phi_c$ and $l_\infty + a$ respectively, where l_∞ is the length of any ray from infinity to the line MM′. The area σ is

$$2(\tfrac{1}{2}a^2 2\phi_c - a^2 \sin\phi_c \cos\phi_c) = a^2(2\phi_c - 2\sin\phi_c\cos\phi_c) = a(2a\phi_c - a\cos\xi).$$

Thus it is verified that the optical paths are equal.

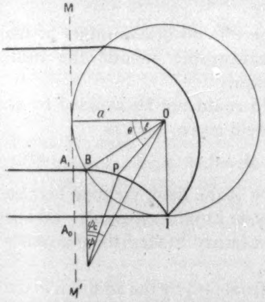


Figure 1.

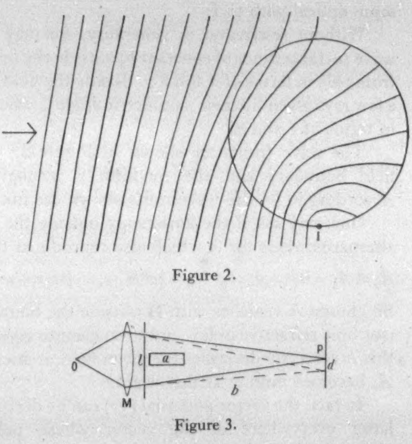


Figure 2.

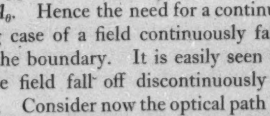


Figure 3.

In order to construct the wave fronts a vector potential describing the given field is required. It is readily seen from the expression for curl **A** in cylindrical coordinates that

$$A_r = A_x = 0; \quad A_\theta = \tfrac{1}{2} r H_0, \quad 0 \leqslant r \leqslant a; \quad A_\theta = \tfrac{1}{2}(a^2 H_0/r), \quad a < r \qquad \ldots\ldots(46)$$

satisfies this condition. If A_θ had the value given in the inner region but vanished in the outer region, an infinite field of opposite sign at the boundary would be indicated, owing to the discontinuity of A_θ. Hence the need for a continuous A_θ. The field discussed here is the limiting case of a field continuously falling off from H_0 to 0 in the neighbourhood of the boundary. It is easily seen that the simplification introduced by letting the field fall off discontinuously has no significant effect on the results given here. Consider now the optical path length p of the ray A_1BP coming from infinity to a point P outside the field, and let P be given by its distance x from MM′ and the angle $\theta = \theta'$.

$$p = \int_{-\infty}^{P} \mu\, ds = \int_{-\infty}^{P} \left(1 + \frac{(\mathbf{AS})}{(H\rho)}\right) ds = l_\infty + x + \frac{1}{H_0 a}\int_{-\infty}^{P} \tfrac{1}{2}\frac{a^2 H_0}{r}\sin\theta\, ds,$$

since **A** is always perpendicular to r the radius vector from 0. With $\sin\theta\, ds = r\, d\theta$ the optical path considered becomes

$$p = l_\infty + x + \tfrac{1}{2}a\theta'. \qquad \ldots\ldots(47)$$

If P is inside the field, the part outside the field is given by (47) with $x=a(1-\cos\xi)$ and $\theta'=\xi$. For the part BP in the field $(\mathbf{AS})=\frac{1}{2}rH_0\sin(\theta+\phi)$; $ds=ad\phi$, so that the optical length of the arc BP is given by $a\phi+\frac{1}{2}\int_0^\varphi r\sin(\theta+\phi)d\phi$. Since the chord BP of length $2a\sin\phi/2$ makes an angle $\phi/2$ with the direction A_1B,

$$r\sin\theta=a\sin\xi+2a\sin^2\phi/2;\qquad r\cos\theta=a\cos\xi-2a\sin\phi/2\,.\,\cos\phi/2,$$

then $r\sin(\theta+\phi)=a\sin(\xi+\phi)-2a\sin^2\phi/2$. Thus

$$\tfrac{1}{2}\int_0^\varphi r\sin(\theta+\phi)d\phi=\tfrac{1}{2}\{-a\cos(\xi+\phi)+a\cos\xi+a\sin\phi-a\phi\},$$

and the total optical path from infinity to the point P is

$$p=l_\infty+a+\tfrac{1}{2}a\{\xi-\cos\xi+\sin\phi+\phi-\cos(\xi+\phi)\}.\qquad\ldots\ldots(48)$$

Notice that if P goes to I, i.e. ϕ goes to $\pi/2-\xi$ this expression becomes $l_\infty+a(1+\pi/4)$ a constant value independent of ξ, confirming that all the rays A_0, A_1 etc. have the same optical path to I.

Without restriction in generality, we may write $l_\infty=0$ and $a=1$, when the wave surfaces (lines of constant p) outside the field can immediately be constructed from (47) in terms of x and θ'. Inside the field, $p(\phi)$ may be plotted from (48) for a few rays given by their parametric angle ξ, when points of given p are constructed in terms of ξ and ϕ.

The wave fronts are shown in Figure 2. There is no discontinuity at the field boundary and the medium is strongly anisotropic outside the field. According to §4 the wavefronts shown are not unique.

One may ask if the anisotropy outside the field could not be avoided by an alternative value for **A** which also reproduces the field given, such as

$$A_r=A_x=0,\qquad A_\theta=\tfrac{1}{2}rH_0-\tfrac{1}{2}a^2H_0/r,\qquad 0\leqslant r\leqslant a;\qquad A_\theta=0,\qquad a>r.\qquad\ldots\ldots(49)$$

So chosen **A** vanishes with **H** outside the boundary of the field, yielding here an isotropic refractive index, and a wavelength given by de Broglie's formula. While this **A** correctly describes the given field, it does not satisfy Stokes' theorem since A_θ becomes infinite at the origin.

In fact, the vector potential (49) can be derived from (46) by the addition to the latter everywhere of the vector whose polar components are $A_2=A_r=0$, $A_\theta=-\frac{1}{2}a^2H_0/r$, that is, of the vector grad ψ where $\psi=-\frac{1}{2}a^2H_0\theta$. The new term would add to the integral $\int_{-\infty}^{1}\mu\,ds$ considered above an amount $-\frac{1}{2}a^2H_0(\theta_\infty-\theta_1)$. This has one value for all rays that pass below the origin O and another for all rays passing above it. If θ is measured as indicated in Figure 1 and ξ_c denotes the ray passing through the origin, the effect of the new term is to add the constant value $a^2H_0\pi/4$ to the integral $\int_{-\infty}^{I}\mu\,ds$ for all rays $\xi_c>\xi>\pi/2$ while it adds $-a^2H_0(3\pi/4)$ for all rays $\xi_c>\xi>-\pi/2$. This means that the oncoming electrons cannot all be described by a single plane wave but that there are two plane waves necessary that cannot be related in phase. This effect is associated with the breakdown of Stokes' theorem for grad θ, the θ component of which is infinite at the origin.

It is readily seen that no vector potential which satisfies Stokes' theorem will remove the anisotropy of the whole space outside the field, so that a consistent optical description appears to be impossible while the whole field-free space is anisotropic. The irremovable anisotropy of the field-free region as a whole emphasizes the fact that the electron-optical refractive index contains the vector potential and not the magnetic field strength. The formulation for the optical path difference given in equation (42) which does not contain **A** does not alter this situation, as a surface integral over H is equivalent to a vector potential. One might therefore expect wave-optical phenomena to arise which are due to the presence of a magnetic field but not due to the magnetic field itself, i.e. which arise whilst the rays are in field-free regions only. Consider now an arrangement as in Figure 3. O denotes a point source of electrons which is focused by the lens M at the point P. Through the pair of slits separated by l a set of interference fringes will arise so that the distance of the nth maximum from P is given by $d = b\lambda_0 n/l$. If, now, a magnetic flux is established normal to the plane of the paper through the area a, then, according to (43), the order of interference at any point of the focal plane is changed by

$$N = \frac{1}{\lambda_0(H\rho)}\iint H_n d\sigma, \quad \text{or with (35) by} \quad N = (\boldsymbol{e}/\boldsymbol{ch})\iint H_n d\sigma. \qquad \ldots\ldots(50)$$

Thus, a flux of $3{\cdot}9 \times 10^{-7}$ gauss cm^2 is required to change the order of interference by 1, and half of this flux will change the maximum at P to a minimum.

It is very curious that equation (50) associates a phenomenon observable at least in principle with a flux; one expects a change of flux, but not steady flux, to have observable effects. The effect has, however, a certain analogy in the existence of a permanent current in a superconducting ring due to a magnetic flux through it.

REFERENCES

Bolza, O., 1909, *Vorlesungen über Variationsrechnung* (Leipzig und Berlin: Teubner).
de Broglie, L., 1930, *An Introduction to the Study of Wave Mechanics* (London: Methuen).
Forsyth, A. R., 1927, *Calculus of Variations* (London: Macmillan).
Frank, P., 1933, *Z. Phys.*, **80**, 4.
Gabor, D., 1945, *Proc. Inst. Radio Engrs.*, **33**, 792.
Glaser, W., 1933, *Z. Phys.*, **80**, 450; 1937, *Beiträge zur Elektronenoptik* (Leipzig: Barth).
Kellogg, O., 1929, *Foundations of Potential Theory* (Berlin: Springer).
Korsunsky, M., 1945, *J. Phys. U.S.S.R.*, **9**, 14.
Maloff, I. G., and Epstein, D. W., 1938, *Electron Optics in Television* (New York: McGraw-Hill).
Opatowski, I., 1943, *Phys. Rev.*, **65**, 54.
Schwarzschild, K., 1903, *Nachr. Ges. Wiss. Göttingen*, **3**, 126.
Sommerfeld, A., und Runge, J., 1911, *Ann. Phys., Lpz.*, **35**, 277.
Whittaker, E. T., 1917, *Analytical Dynamics* (Cambridge: University Press).
Zworykin, V. K., Morton, G. A., Ramberg, E. G., Hillier, J., Vance, A. W., 1945, *Electron Optics and the Electron Microscope* (New York: John Wiley & Sons).

The Disturbance near the Focus of Waves of Radially Non-uniform Amplitude

By H. H. HOPKINS

I.C.I. Research Fellow, Technical Optics Department, Imperial College, London

MS. received 21*st February* 1947, *in revised form* 14*th July* 1948; *read before the Optical Group* 3*rd October* 1947

ABSTRACT. Lommel's problem, the disturbance near the focus of waves of uniform amplitude, is generalized to consider waves of radially non-uniform amplitude. Lommel's numerical results are extended, and his *U* and *V* functions, tabulated on a new basis, together with some new related functions, are employed to provide extensive intensity distribution curves for the diffracting patterns associated with waves of both uniform and non-uniform amplitude.

---

§ 1. INTRODUCTION

THE distribution of intensity in the geometrical focal plane of a spherical wave was derived by Airy (1834). It has the form

$$I = \{2J_1(z)/z\}^2, \qquad \ldots\ldots(1)$$

where I is the intensity at points distance ρ' from the centre of the diffraction pattern, and

$$z = 2\pi r \rho'/\lambda R. \qquad \ldots\ldots(2)$$

The radius of the circular aperture which limits the wave is r, and R is the distance between the limiting aperture and the focal plane. λ is the wavelength of the disturbance in the medium in question.

The expression (1) accounts for the form of the diffraction pattern, consisting of a bright central maximum surrounded by alternate dark and bright rings, which is observed as the image of a distant star in a well-corrected telescope. The dependence of the scale of the diffraction pattern on the relative aperture and the wavelength is shown by (2).

In deriving (1) it is assumed that the angular semi-aperture is small, and that the wave is of uniform amplitude. The former restriction has been removed in an earlier communication (Hopkins 1943), the error in the approximate formula being shown not to exceed a few per cent for $\alpha < 30°$, where α is the angular semi-aperture.

With the approximation implied in Airy's formula, Lommel (1886) considered the intensity distributions in planes other than the geometrical focal plane. Development in series of the integrals concerned led to the introduction of the Lommel *U* and *V* functions.

It is proposed here to remove the restriction of uniform amplitude from Lommel's problem, and to investigate formulae expressing the disturbance near the focus of a spherical wave showing a radial non-uniformity of amplitude. It is found necessary to introduce both the Lommel functions and new related functions. These functions have been tabulated and used to provide extensive numerical results.

§2. LOMMEL'S PROBLEM FOR WAVES OF RADIALLY NON-UNIFORM AMPLITUDE

In Figure 1, O(X, Y, Z) is a system of rectangular coordinates, O being the centre of an approaching spherical wave. P (X, Y, Z) is a point on this latter, and YOZ is the geometrical focal plane. The wave is limited by a circular aperture of radius r, and $OP = R$ is the radius of curvature of the wave.

For small angular apertures one can write $\rho\, d\rho\, d\phi$ for the area of an element of wave at P, where (ρ, ϕ) are polar coordinates of P, having origin on OX and initial line in the plane ZOX. The disturbance at $P'(\Delta X, Y', Z')$, a point in a plane of focus a small distance ΔX from YOZ, can be written

$$s = \int_0^r \int_0^{2\pi} A \sin 2\pi(t/T - d/\lambda)\rho\, d\rho\, d\phi, \qquad \ldots\ldots(3)$$

where $d = PP'$, A is the amplitude at P, T is the period of the disturbance, and t the time variable.

The distance d is given by $d^2 = (X - \Delta X)^2 + (Y - Y')^2 + (Z - Z')^2$ or, ΔX, Y' and Z' being small, $d^2 = R^2 - 2\rho\rho' \cos(\phi - \phi') - 2X\Delta X$, where (Y, Z), (Y', Z') have been changed to polar coordinates (ρ, ϕ), (ρ', ϕ'), the latter referring to the point P′ in the plane of focus through O′. The coordinate X is given by $X = (R^2 - \rho^2)^{\frac{1}{2}} = R - (\rho^2/2R)$ with the required accuracy. Thus

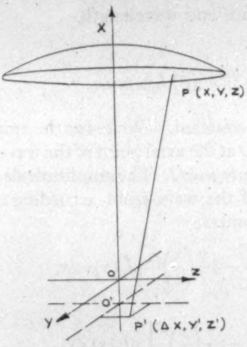


Figure 1.

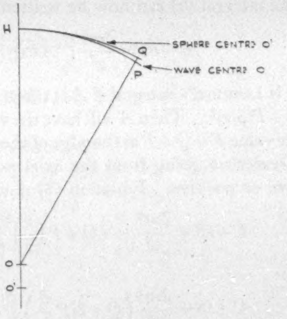


Figure 2.

$$d = (R - \Delta X) - (\rho\rho'/R)\cos(\phi - \phi') + (\rho^2/2R^2)\Delta X. \qquad \ldots\ldots(4)$$

Substitution of (4) in (3), and the expansion of the argument of the sine in the integrand in terms of those parts dependent upon, and those independent of, the variables ρ, ϕ gives

$$s = \sqrt{(C^2 + S^2)} \sin 2\pi(t/T - (R - \Delta X)/\lambda + \epsilon),$$

where $\tan \epsilon = C/S$ and

$$\begin{matrix} C \\ S \end{matrix} = \int_0^r \int_0^{2\pi} A \begin{matrix} \cos \\ \sin \end{matrix} \left\{ \frac{2\pi\rho\rho'}{\lambda R} \cos(\phi - \phi') - \frac{\pi\rho^2}{\lambda R^2}\Delta X \right\} \rho\, d\rho\, d\phi.$$

Thus the intensity at P′ can be expressed as the squared modulus of the integral

$$C+iS=\int_0^r\int_0^{2\pi} A\exp i\{(2\pi\rho\rho'/\lambda R)\cos(\phi-\phi')-(\pi\rho^2/\lambda R^2)\Delta X\}\rho\,d\rho\,d\phi. \quad \ldots\ldots(5)$$

If now the amplitude A varies radially, one may write (5)

$$C+iS=\int_0^r A\exp\{-i(\pi\rho^2/\lambda R^2)\Delta X\}\rho\,d\rho\int_0^{2\pi}\exp\{i(2\pi\rho\rho'/\lambda R)\cos(\phi-\phi')\}d\phi$$

$$=2\pi\int_0^r A\exp(-iy/2)(\rho/r)^2J_0(x)\rho\,d\rho, \quad \ldots\ldots(6)$$

where $y/2=(\pi r^2/\lambda R^2)\Delta X$, and $x=(2\pi/\lambda)(\rho\rho'/R)$. The significance of these will be seen from the following. In Figure 2, HP is a section through the spherical wave-front, and HQ a section through a sphere centred on O′. The marginal difference PQ measured in wavelengths, is

$$W=(\mathrm{PQ})/\lambda=r^2\Delta X/2\lambda R^2, \qquad \text{or} \qquad y/2=2\pi W. \quad \ldots\ldots(7)$$

It denotes the amount out-of-focus, independent of the aperture.

Again, let z be the value of x when $\rho=r$, then $\rho=rx/z$; $\rho\,d\rho=(r^2/z^2)x\,dx$; and $z=2\pi r\rho'/\lambda R$, which is the variable defined in (2). It denotes the scale of the diffraction pattern independent of the aperture and wavelength.

The integral (6) can now be written

$$C+iS=\frac{2\pi r^2}{z^2}\int_0^z A\exp[-i(y/2)(x/z)^2]J_0(x)x\,dx, \quad \ldots\ldots(8)$$

which is Lommel's integral if A is taken to be constant. Write for the amplitude $A=Q+P(\rho/r)^2$. Then A will have the value Q at the axial point of the wave-front, and the value $F=Q+P$ at the edge of the aperture $\rho=r$. The amplitude decreases or increases on going from the axial point of the wave-front according as P is negative or positive. Equation (8) now becomes

$$C+iS=\frac{2\pi r^2}{z^2}\int_0^z\left\{Q+P\left(\frac{x}{z}\right)^2\right\}\exp\left[-i\frac{y}{2}\left(\frac{x}{z}\right)^2\right]J_0(x)x\,dx,$$

or

$$C+iS=\frac{2\pi r^2}{z^2}\left(Q+2iP\frac{\partial}{\partial y}\right)\int_0^z\exp\left[-i\frac{y}{2}\left(\frac{x}{z}\right)^2\right]xJ_0(x)\,dx. \quad \ldots\ldots(9)$$

The integral to be evaluated in now simply that of Lommel's original problem.* This is expressed in two forms, suited respectively to the regions $z\leqslant y$, $z\geqslant y$.

These results are (Gray and Mathews 1922)

$$\frac{2}{z^2}\int_0^z\exp\left[-i\frac{y}{2}\left(\frac{x}{z}\right)^2\right]xJ_0(x)\,dx=\frac{2}{y}\left(\sin\frac{z^2}{2y}+V_0\sin\tfrac{1}{2}y-V_1\cos\tfrac{1}{2}y\right)$$

$$+i\frac{2}{y}\left(\cos\frac{z^2}{2y}-V_0\cos\tfrac{1}{2}y-V_1\sin\tfrac{1}{2}y\right), \quad \ldots\ldots(10)$$

for the region $z\leqslant y$, and

$$\frac{2}{z^2}\int_0^z\exp\left[-i\frac{y}{2}\left(\frac{x}{z}\right)^2\right]xJ_0(x)\,dx=\frac{2}{y}(U_1\cos\tfrac{1}{2}y+U_2\sin\tfrac{1}{2}y)$$

$$+i\frac{2}{y}(U_1\sin\tfrac{1}{2}y-U_2\cos\tfrac{1}{2}y) \quad \ldots\ldots(11)$$

---

* This step shortens the analysis. I am indebted to a referee of the paper for pointing it out.

for the region $z \geqslant y$, where

$$U_n = \sum_{p=0}^{\infty} (-1)^p (y/z)^{n+2p} J_{n+2p}(z); \quad V_n = \sum_{p=0}^{\infty} (-1)^p (z/y)^{n+2p} J_{n+2p}(z). \quad \ldots\ldots(12)$$

The results (10) and (11) immediately yield expressions for (9) suited to these two regions. Thus, omitting the constant factor πr^2, we have, for the region $z \leqslant y$,

$$C+iS = \left(Q+2iP\frac{\partial}{\partial y}\right)\frac{2}{y}\left(\sin\frac{z^2}{2y}+V_0\sin\tfrac{1}{2}y-V_1\cos\tfrac{1}{2}y\right)$$
$$+i\left(Q+2iP\frac{\partial}{\partial y}\right)\frac{2}{y}\left(\cos\frac{z^2}{2y}-V_0\cos\tfrac{1}{2}y-V_1\sin\tfrac{1}{2}y\right),$$

or, with some arrangement of terms,

$$C = F\Gamma + P(2/y)(\Gamma_{\mathrm{A}}-\Sigma); \quad S = F\Sigma + P(2/y)(\Gamma+\Sigma_{\mathrm{A}}), \quad \ldots\ldots(13)$$

where

$$\left.\begin{aligned} \Gamma &= \frac{2}{y}\left(\sin\frac{z^2}{2y}+V_0\sin\tfrac{1}{2}y-V_1\cos\tfrac{1}{2}y\right),\\ \Sigma &= \frac{2}{y}\left(\cos\frac{z^2}{2y}-V_0\cos\tfrac{1}{2}y-V_1\sin\tfrac{1}{2}y\right), \end{aligned}\right\} \quad \ldots\ldots(14)$$

and

$$\left.\begin{aligned} \Gamma_{\mathrm{A}} &= \frac{2}{y}\left\{\left(\frac{z^2}{2y}-\frac{y}{2}\right)\sin\frac{z^2}{2y}+Y_0\cos\tfrac{1}{2}y+Y_1\sin\tfrac{1}{2}y\right\},\\ \Sigma_{\mathrm{A}} &= \frac{2}{y}\left\{\left(\frac{z^2}{2y}-\frac{y}{2}\right)\cos\frac{z^2}{2y}+Y_0\sin\tfrac{1}{2}y-Y_1\cos\tfrac{1}{2}y\right\}. \end{aligned}\right\} \quad \ldots\ldots(15)$$

The Y functions and the X functions are defined by

$$\left.\begin{aligned} X_n &= \frac{1}{y}\frac{\partial U_n}{\partial y} = \sum_{p=0}^{\infty}(-1)^p(n+2p)\left(\frac{y}{z}\right)^{n+2p} J_{n+2p}(z),\\ Y_n &= -y\frac{\partial V_n}{\partial y} = \sum_{p=0}^{\infty}(-1)^p(n+2p)\left(\frac{z}{y}\right)^{n+2p} J_{n+2p}z). \end{aligned}\right\} \quad \ldots\ldots(16)$$

For the region $z \geqslant y$

$$C+iS = \left(Q+2iP\frac{\partial}{\partial y}\right)\frac{2}{y}(U_1\cos\tfrac{1}{2}y+U_2\sin\tfrac{1}{2}y)$$
$$+i\left(Q+2iP\frac{\partial}{\partial y}\right)\frac{2}{y}(U_1\sin\tfrac{1}{2}y-U_2\cos\tfrac{1}{2}y);$$

expressing C and S in the form (13), one must now write

$$\Gamma = (2/y)(U_1\cos\tfrac{1}{2}y+U_2\sin\tfrac{1}{2}y); \quad \Sigma = (2/y)(U_1\sin\tfrac{1}{2}y-U_2\cos\tfrac{1}{2}y) \quad \ldots\ldots(17)$$

and

$$\Gamma_{\mathrm{A}} = (2/y)(X_1\sin\tfrac{1}{2}y-X_2\cos\tfrac{1}{2}y); \quad \Sigma_{\mathrm{A}} = (2/y)(-X_1\cos\tfrac{1}{2}y-X_2\sin\tfrac{1}{2}y). \quad \ldots\ldots(18)$$

It will be seen that in the case $F=Q$ (i.e. $P=0$) the above formulae degenerate to Lommel's results for waves of uniform amplitude. Further, the expressions (13), together with (14), (15), or (17), (18), provide a solution of the present problem, for the intensity is given by $I(y,z)=C^2+S^2$, and can be calculated at once given tables of the U, V, X and Y functions.

§3. GENERAL FEATURES OF THE INTENSITY DISTRIBUTIONS

As one would expect, the intensity is independent of ϕ'. If y be replaced by $-y$, the expression $I = C^2 + S^2$ remains unchanged, although C changes sign. Thus the intensity distributions, considered as functions of z, are identical in planes equidistant from the plane $y = 0$. Hence, as in the case of waves of uniform amplitude, the spatial distribution of intensity is symmetrical both about the geometrical focal plane and about the central line $z = 0$.

(i) *The Geometrical Focal Plane*

Putting $y = 0$, one obtains for the geometrical focal plane

$$C = F[2J_1(z)/z] - P[4J_2(z)/z^2], \qquad S = 0.$$

The intensity is thus

$$I(0, z) = \{F[2J_1(z)/z] - P[4J_2(z)/z^2]\}^2,$$

which yields the Airy formula (1) for $F = Q = 1$, $P = 0$. The intensity at the geometrical focus is $I(0, 0) = (Q + \tfrac{1}{2}P)^2$ in the general case.

(ii) *The Central Line $z = 0$*

Taking the limits of (14), (15) as $z \to 0$, one has

$$\left.\begin{aligned} C &= F(2/y)\sin\tfrac{1}{2}y + P(2/y)^2(\cos\tfrac{1}{2}y - 1), \\ S &= F(2/y)(1 - \cos\tfrac{1}{2}y) + P(2/y)^2(\sin\tfrac{1}{2}y) - 2P/y. \end{aligned}\right\} \qquad \ldots\ldots(19)$$

A zero of intensity requires that C, S be simultaneously zero. That is, since $F = Q + P$,

$$F\sin\tfrac{1}{2}y + (2P/y)\cos\tfrac{1}{2}y = 2P/y; \qquad -F\cos\tfrac{1}{2}y + (2P/y)\sin\tfrac{1}{2}y = -Q,$$

or

$$[F^2 + (2P/y)^2]^{\frac{1}{2}} \begin{matrix}\sin \\ \cos\end{matrix}(\tfrac{1}{2}y + \epsilon) = \begin{matrix}2P/y \\ Q\end{matrix}.$$

Squaring, and adding these equations, $F = Q$. Thus there are no zeros of intensity along the central line, except in the case of waves of uniform amplitude.

Writing $F = Q + P$ in (19), we find for the intensity along the central line

$$I(y, 0) = (Q + \tfrac{1}{2}P)^2\left(\frac{\sin\frac{1}{4}y}{\frac{1}{4}y}\right)^2 + (\tfrac{1}{2}P)^2\left(\frac{4}{y}\right)^2\left(\frac{\sin\frac{1}{4}y}{\frac{1}{4}y} - \cos\tfrac{1}{4}y\right), \qquad \ldots\ldots(20)$$

which is easily interpreted since $\tfrac{1}{4}y = \pi W$ as in (7).

If $Q = F = 1$, $P = 0$, (20) gives $I(y, 0) = (\sin\tfrac{1}{4}y/\tfrac{1}{4}y)^2$ as the intensity distribution for waves of uniform amplitude. The zeros occur at integral values of W, excluding $W = 0$; the maxima occur at the roots of $\tan\pi W = \pi W$. The maxima of (20) again appear at the roots of $\tan\pi W = \pi W$, and the minima at the roots of

$$2Q(Q + P)(\pi W)^2\tan\pi W - P^2(\pi W - \tan\pi W) = 0.$$

The minima are, in fact, displaced towards the positions of the maxima—degenerating, as the above equation shows, to points of inflection for the cases $Q = 0$ and $F = Q + P = 0$ respectively.

At points along the central line corresponding to integral values of W, the intensity is zero for waves of uniform amplitude ($P = 0$). For non-uniform waves the intensity at these points is $I(4\pi W, 0) = (P/2\pi W)^2$.

Thus, at the first "dark point" the intensities, in different cases, are as follows:

$$P=-1{\cdot}0: \quad I(4\pi,0)=0{\cdot}025, \quad I(0,0)=0{\cdot}250,$$
$$P=-0{\cdot}4: \quad I(4\pi,0)=0{\cdot}004, \quad I(0,0)=0{\cdot}640.$$

The first of these cases is that of a wave whose amplitude decreases from 1·0 to zero on going from the centre to the edge of the wave. The intensity at $y=4\pi$ is 10% of that at the geometrical focus. The second case is that of a microscope objective, for which the non-uniformity of amplitude has been calculated using Fresnel's formulae in a manner shown earlier (Hopkins 1944). At the first "dark point" the intensity is 0·6% of that at the geometrical focus.

Consider now two waves, in one of which the central and marginal amplitudes are $Q=\alpha$, $F=\beta$; and in the other $Q=\beta$, $F=\alpha$. In each case (20) becomes

$$I=(y,0)=\left(\frac{\alpha+\beta}{2}\right)^2\left(\frac{\sin\pi W}{\pi W}\right)^2+\left(\frac{-\beta}{2}\right)^2\left(\frac{1}{\pi W}\right)^2\left(\frac{\sin\pi W}{\pi W}-\cos\pi W\right),$$

the interesting result that the intensity distributions along the central line are the same. However, considered as functions of z, the two cases differ widely.

§4. TABULATION OF THE U, V AND X FUNCTIONS

Lommel has provided values of the functions U_1, U_2, V_0, V_1 for $y=\pi$, $2\pi, \dots.\ 10\pi$, and $z=0, 1, \dots.\ 12$. For problems in the diffraction theory of spherical aberration, using a method developed earlier (Hopkins 1945), it is desirable to cover the ranges $y=-\infty$ to $+\infty$ and $z=0, 1, \dots.\ 36$. The following method of tabulation was therefore adopted in each case. Writing $A=y/z$, and $B=z/y$, the definitions (12) and (16) give

$$\begin{matrix}U\\V\end{matrix}=\sum_{p=0}^{\infty}(-1)^p\begin{pmatrix}A\\B\end{pmatrix}^{n+2p}J_{n+2p}(z); \qquad \begin{matrix}X\\Y\end{matrix}=\sum_{p=0}^{\infty}(-1)^p\begin{pmatrix}A\\B\end{pmatrix}^{n+2p}(n+2p)J_{n+2p}(z),$$

that is, for $A=B$, $U_n=V_n$, $X_n=Y_n$.

For even values of n, the functions do not change sign with the argument A or B, while the sign is reversed for odd values of n. A table showing U_n, X_n for $A=0, 0{\cdot}05, \dots.\ 1{\cdot}0$ and $z=0, 1, \dots.\ 36$ will thus cover U_n, V_n, X_n and Y_n for the required ranges.

The tables were computed by machine, using five-figure values taken from the tables of Bessel functions given by Gray and Mathews (1922).

The functions reduce to simple expressions when $A=0$ or $A=1$. These were used to check the terminal values, intermediate ones being checked by plotting the first differences as functions of A. Three places of decimals were found to be quite adequate for the present problem.

§5. INTENSITY DISTRIBUTION CURVES

In the case of a light wave originating from an axial point source, the energy transmitted along marginal ray-paths through a lens system may differ greatly from that along the axis, since here the incidence is always normal, whereas along marginal rays large angles of incidence may be encountered. It is not difficult to show that the variation of amplitude with aperture should be a linear function of the square of the aperture. This is the case treated above.

In an F/2 photographic objective, for example, the amplitude at the margin was but 70% of that at the centre, while in the case of the microscope objective referred to above it was but 60%. Thus it seems reasonable to assume that the case $Q=1{\cdot}0$, $F=0{\cdot}5$ would represent the very extreme likely to be met with in practice. On the other hand, an artificial "inhomogenizing" of amplitude is limited at the cases $Q=0{\cdot}0$, $F=1{\cdot}0$ and $Q=1{\cdot}0$, $F=0{\cdot}0$. These cases have therefore been computed in order to investigate the importance of the effect in common practice, and its theoretical limits. The curves are all shown on a scale such that the intensity is unity at the geometrical focus.

(i) *Uniform Amplitude.* $Q=F=1$

In each figure the intensity curve for this case is shown as a broken line. In each plane of focus the intensity is computed for $z=0, 1, \ldots . 24$, thus extending Lommel's results, which are given for values of z up to 12.

It will be seen that the out-of-focus distributions approach the "Airy disc" (shown as the broken line) for large values of z. It is not difficult to show this to be required by the above formulae.

(ii) *The Central Line* $z=0$

The intensity curves for three cases are shown in figure 3. The marked features that appear as the non-uniformity of amplitude increase are the disappearance of the "dark points" and the increase in depth of focus. This latter is further exemplified in the curves of (iv).

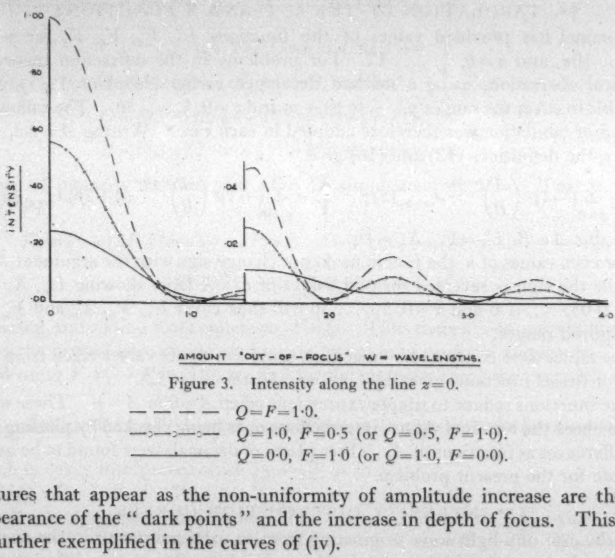


Figure 3. Intensity along the line $z=0$.

— — — — $Q=F=1{\cdot}0$.

—·—·—·— $Q=1{\cdot}0$, $F=0{\cdot}5$ (or $Q=0{\cdot}5$, $F=1{\cdot}0$).

———— $Q=0{\cdot}0$, $F=1{\cdot}0$ (or $Q=1{\cdot}0$, $F=0{\cdot}0$).

(iii) *Non-uniform Amplitude.* $Q=1{\cdot}0$, $F=0{\cdot}5$

The intensity curves in this case, those of a wave whose amplitude decreases to a value at the edge equal to half that at the centre, are shown in Figures 4 to 12. It will be seen that the differences between the forms of the light distributions curves of this case and those of a wave of uniform amplitude are not very significant from a practical point of view. It is therefore justifiable, in the case of almost every actual lens system, to assume the emergent wave-fronts to be of uniform amplitude so far as the form of the light distribution curves are concerned.

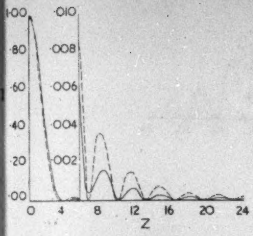


Figure 4.

Plane of focus : $W=0$.

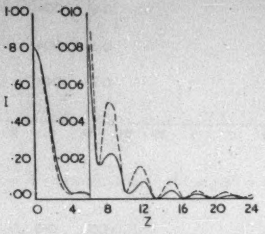


Figure 5.

Plane of focus : $W=\frac{1}{4}$.

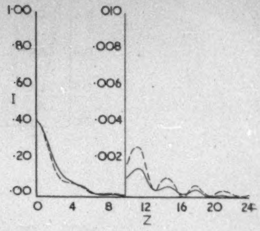


Figure 6.

Plane of focus : $W=\frac{1}{2}$.

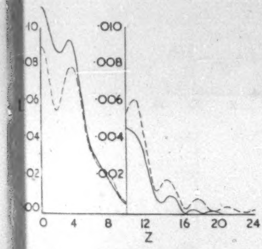


Figure 7.

Plane of focus : $W=\frac{3}{4}$

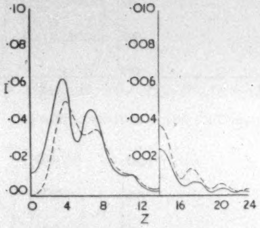


Figure 8.

Plane of focus : $W=1$.

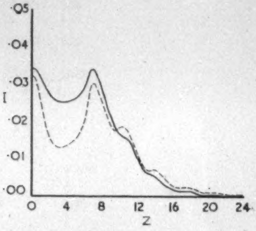


Figure 9.

Plane of focus : $W=1\frac{1}{4}$.

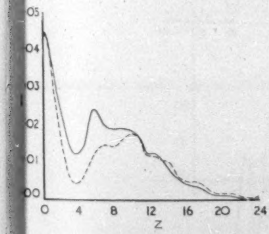


Figure 10.

Plane of focus : $W=1\frac{1}{2}$.

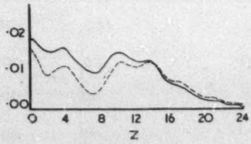


Figure 11.

Plane of focus : $W=1\frac{3}{4}$.

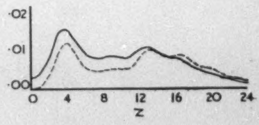


Figure 12.

Plane of focus : $W=2$.

– – – $Q=F=1$. ——— $Q=1$, $F=\frac{1}{2}$ (on 1·7775 × scale).

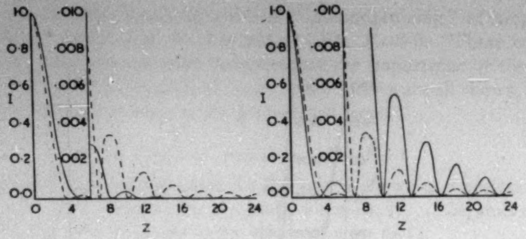


Figure 13. Plane of focus: $W=0$.

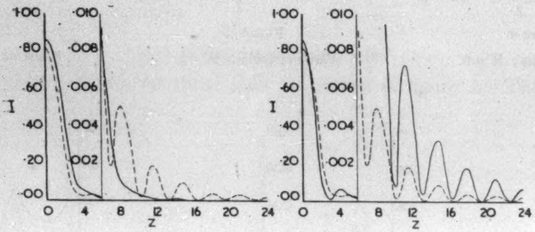


Figure 14. Plane of focus: $W=\frac{1}{4}$.

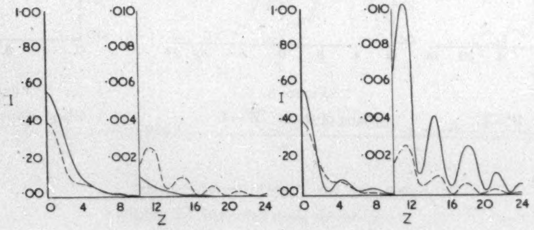


Figure 15. Plane of focus: $W=\frac{1}{2}$.

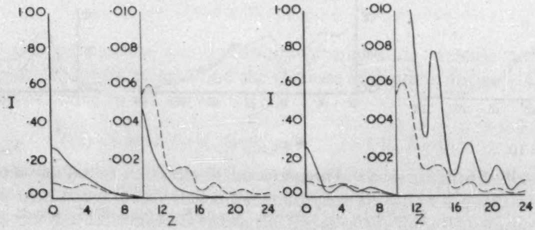


Figure 16. Plane of focus: $W=\frac{3}{4}$.

– – – $Q=F=1$. – – – $Q=F=1$.
——— $Q=1$, $F=0$ (4 × scale). ——— $Q=0$, $F=1$ (4 × scale).

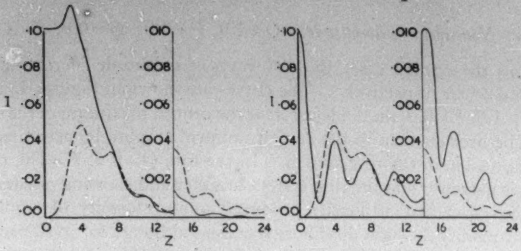


Figure 17. Plane of focus: $W=1$.

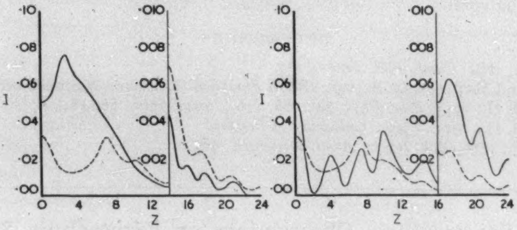


Figure 18. Plane of focus: $W=1\frac{1}{4}$.

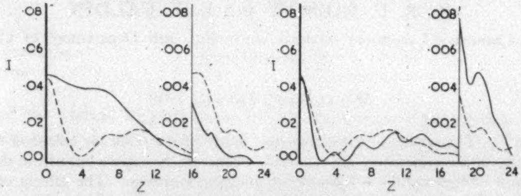


Figure 19. Plane of focus: $W=1\frac{1}{2}$.

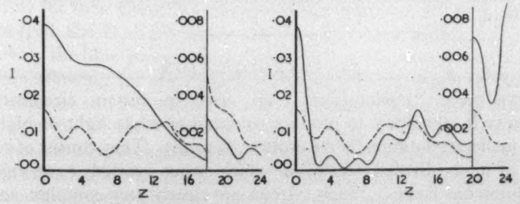


Figure 20. Plane of focus: $W=1\frac{3}{4}$.

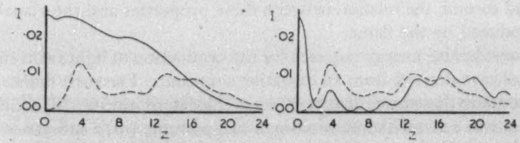


Figure 21. Plane of focus: $W=2$.

– – – $Q=F=1$.
——— $Q=1$, $F=0$ (4× scale).

– – – $Q=F=1$.
——— $Q=0$, $F=1$ (4× scale).

(iv) *Non-uniform amplitude.* $Q=1\cdot0$, $F=0\cdot0$; $Q=0\cdot0$, $F=1\cdot0$

These are the extreme cases, those of waves of amplitude decreasing to zero at the edge and centre respectively. The curves are shown in Figures 13 to 21. In the case $Q=1\cdot0$, $F=0\cdot0$, the tendency is for the central maximum of the diffraction pattern to be broader than in the case of uniform amplitude, accompanied by a virtual suppression of the ring system. In the case $Q=0\cdot0$, $F=1\cdot0$, the central maximum is narrower and the ring system brighter and showing greater contrast.

It is thus seen that artificially produced non-uniformity of amplitude can significantly modify the light distribution curves, although for most ordinary lens systems it is quite justifiable to assume uniform amplitude in evaluating the diffraction integral.

REFERENCES

AIRY, G. B., 1834, *Camb. Phil. Trans.*, 283.
GRAY, A., and MATHEWS, G. B., 1922, *Bessel Functions* (London : Macmillan & Co.).
HOPKINS, H. H., 1943, *Proc. Phys. Soc.*, **55**, 116 ; 1944, *Ibid.*, **56**, 48.
HOPKINS, H. H., 1945, *Thesis, University of London.*
LOMMEL, W., 1886, *Abh. Bayer. Akad. Wissensch.*, **15.**

Some Spectroscopic Observations on Pyrotechnic Flames

BY R. F. BARROW AND E. F. CALDIN

Physical Chemistry Laboratory, Oxford University, and Department of Chemistry,
Leeds University

MS. received 3 *February* 1948

ABSTRACT. The results of a spectroscopic study of the white or coloured flames from typical pyrotechnic compositions are reported. Most of the visible light from these sources comes from a limited number of atomic or diatomic emitters. The effects of secondary emitters on the dominant wavelength, colorimetric purity and relative intensity are considered in general terms and some suggestions for improving the quality of the flames are put forward.

§ 1. INTRODUCTION

PYROTECHNIC "compositions" are used in certain circumstances, for instance in signalling, to produce coloured or white light of high intensity and moderate duration (of the order of seconds). They consist of compressed mixtures of finely divided substances which, when ignited, react vigorously to produce luminous flames. Such systems are clearly very complex, and we shall consider only two main aspects: first, the spectroscopic characteristics of the flames, and second, the relation between these properties and the visual sensation that is produced by the flame.

The considerable energy required for the production of light from such a flame is almost always derived from an oxidation reaction. For such flames, the composition contains (i) a source of oxygen, e.g. a chlorate or nitrate; (ii) a fuel, such as a carbohydrate or a metal like aluminium or magnesium; (iii) a substance which will impart colour to the flame, e.g. cuprous chloride; and possibly (iv) small amounts of substances to protect an electro-positive metal, to produce gas to increase the size of the flame, to act as a binder for the compressed composition, or to modify the rate of burning. Sometimes, of course, it is possible to choose a substance to fulfil simultaneously more than one of these various functions. These general remarks are illustrated by some specimen compositions given in the next section.

§ 2. SPECTROSCOPIC PROPERTIES

The spectra of the flames were photographed in a first order of a 2·4 m. concave grating in an Eagle mounting, which gives a nearly linear dispersion of about 7·4 A/mm. Ilford "Special Long Range Panchromatic" plates were used. The spectra were normally examined over the range 3000 to 8000 A.

Altogether a large number of different compositions were studied: those listed below (taken from Weingart 1943 and Davis 1943) are typical examples.

Blue	%	*Green*	%
Potassium chlorate	63	Barium nitrate	50
Copper carbonate	15	Barium chlorate	40
Mercurous chloride	10	Shellac	10
Shellac	12		
Yellow		*Red*	
Potassium chlorate	55	Potassium chlorate	15
Sodium oxalate	30	Strontium nitrate	65
Shellac	15	Shellac	20
White		*White*	
Aluminium	22	Magnesium	50
Barium nitrate	75	Barium nitrate	45
Sulphur	3	Paraffin wax	5

Most of the visible light from these sources comes from various atomic or diatomic emitters. It will be best to consider the different types of composition in turn.

Blue flames.

These rely for their colour on emission of the spectrum of CuCl. Most of the light comes from the D and E systems in the range 4200–4600 A. (cf. Pearse and Gaydon 1941); the blue-green systems B and C, and the green system A, are much weaker (even allowing for the drop in sensitivity of the plates in this region). Potassium chlorate and perchlorate are often used both as oxidizing agents and as sources of chlorine for the formation of CuCl in the flame; this leads to strong emission from the K resonance lines at 7665 and 7699 A. The blue composition given above contains mercurous chloride, which, however, can have no function except to provide chlorine (Wieland 1945); equally satisfactory compositions may be prepared which contain none of this substance.

Green flames.

BaCl, which gives rise to a well-known band system at 5050–5350 A., is the obvious choice for a green emitter, and it is in fact generally used. The spectrograms of such flames usually show, in addition to this system, band-systems of CaCl and of SrCl in the orange and red regions of the spectrum, which evidently arise from small amounts of impurities in the original barium compounds, and they invariably show the extensive band-system of BaO (4000–8000 A.). There is also strong emission from a near infra-red system of BaCl (Barrow and Crawford 1946) which, however, lies beyond the visible region of the spectrum. Other discrete bands in the region 7000–7600 A. have not yet been identified.

Yellow flames.

When sodium is introduced into a flame such as that of hydrogen burning in air, the emission spectrum is found to depend markedly on the concentration of sodium. Even at quite low concentrations (of the order of 0·001 gm/litre of hydrogen in a flame which consumes 1·5 litres/min. of hydrogen) the D lines are strongly reversed, and continuous emission extending for several hundreds of Ångströms on either side of the D lines is developed. At high concentrations, such as may be obtained by burning sodium metal in an atmosphere of air enriched with oxygen, there is strong emission from the continuum throughout the range 5500 to 7000 A. (Crawford 1944, Hamada 1933). The visible radiation from all the yellow flames examined consisted primarily of emission from the D lines and from the associated continuum.

Red flames.

Strontium compounds are used to colour the red flames, but there are still some obscure features in the spectra. The simultaneous presence of strontium and chlorine compounds in the compositions leads to strong emission of the red system of SrCl (the violet system also appears weakly). However, there are also present two closely-spaced sequences degraded to the violet which do not belong to the SrCl system. The long wavelength edges of these sequences are at 6884·5 and 6114·2 A. In strontium flames containing no chlorine compound these heads are more strongly developed, and, in addition, the region between them can be seen to be filled with a very close pattern of what appear to be rotational structure lines. The most likely emitter would seem to be SrO, and it is some support for this suggestion that a heavy-current positive-column discharge through SrO contained in a silica tube leads to the production of the same bands.

White flames.

Some of the possible ways of making "white" flames will be obvious from what has gone before. Three methods which have been used are: (i) to develop an extensive continuum, e.g. the sodium continuum, in emission, (ii) to excite an extensive discrete band-system, and (iii) to excite simultaneously two more or less complementary band-systems. The most suitable system for (ii) appears to be that of BaO, which may be produced conveniently by the oxidation of aluminium or magnesium by barium nitrate, while for (iii) suitable blending of the emission from SrCl (red), CaCl (yellow) and BaCl (green) has been found to yield an almost white light.

§ 3. COLOUR AND SPECTRAL DISTRIBUTION OF ENERGY

The spectral distribution of energy in the light from these flames has not been quantitatively determined, but the effects of light from secondary emitters on the colours of the flames can be considered in general terms. The red flames depending on the emission from SrCl and SrO emit also yellow light (due to Na), green (due to BaCl) and blue (due to Sr). Green flames may emit, besides the green light from BaCl, some yellow from Na, red from SrCl and nearly white light from BaO. Blue flames, which rely on the D and E band systems of CuCl, emit also green light (from the A system of CuCl), yellow (from Na) and red (from K). Calculations have been made on the effects of "diluting" light of each of the relevant colours with light of the wavelengths due to secondary emitters. Actually seven wavelengths have been considered and computations made for mixtures of light of each pair of wavelengths in proportions covering the whole scale from 1 : 0 to 0 : 1. These seven wavelengths cover both the main and the secondary emitters (see Table 1).

Table 1

Colour	Wavelength A.	Corresponding emitters with wavelength in A.
Red I	6900	SrCl, *ca.* 6200–6800; SrO, *ca.* 6000–6900;
Red II	6200	CaCl, 5800–6380; K, 7699 and 7665.
Yellow	5890	Na, 5890, 5896.
Green I	5300	BaCl, 5050–5350.
Green II	5100	CuCl, A system, *ca.* 5100–5500.
Blue I	4600	Sr, 4607; CuCl, D and E systems, 4200–4600.
Blue II	4100	

The method of computation is as follows. From the C.I.E. trichromatic coefficients for spectral light of given wavelengths, the corresponding specification for a series of mixtures is found (Wright 1944, Moon 1938). Let the trichromatic coefficients of two spectral colours to be mixed be (x_1, y_1, z_1), (x_2, y_2, z_2), and those of the resulting mixture be (x_m, y_m, z_m). Then if α_1 energy units of the first spectral colour be mixed with α_2 energy units of the second (where $\alpha_1 + \alpha_2 = 1$), the trichromatic coefficients of the mixture will be given by $x_m = \alpha_1 x_1 + \alpha_2 x_2$, $y_m = \alpha_1 y_1 + \alpha_2 y_2$.

Having found the trichromatic coefficients for a given mixture in this way, one finds the dominant wavelength with the aid of the C.I.E. diagram. A straight line is drawn from the point on the diagram representing equal-energy white light ($x_w = 0{\cdot}333$, $y_w = 0{\cdot}333$) to the point representing the mixture (x_m, y_m) and produced to meet the locus of spectral colours (or, for purples, the line joining the extremities of the locus). Let the point of intersection be (x_s, y_s). The dominant wavelength of the mixture is given by the value of λ corresponding to (x_s, y_s), the relative luminous intensity, by the value of y_m, and the colorimetric purity by the formulae (Hardy 1936)

$$P = \frac{y_s}{y_m} \frac{y_m - y_w}{y_s - y_w} = \frac{y_s}{y_m} \frac{x_m - x_w}{x_s - x_w}.$$

The dominant wavelength, colorimetric purity, and relative intensity correspond roughly with the sensation attributes hue, saturation, and brilliancy, and give a fair idea of the appearance of a coloured light. The results are given in Table 2.

Certain general deductions can be made from these figures, even though quantitative spectrophotometric data are wanting. (The variations of the several characteristics with the proportions of two colours are best seen by plotting them on graphs). As regards hue, the colours most easily spoiled by addition of light of other colours seem to be red and blue; as regards saturation, blue is much the most easily spoiled; as regards brilliancy, yellow and green. Blue light also causes the greatest decrease in purity when added to light of other colours. Addition of white light will reduce purity without altering dominant wavelength; the emission of the BaO spectrum will have some such effect on green flames.

Table 2

Percentage of total energy as light of first-named colour	100	80	60	40	20	0	100	80	60	40	20	0
Colour pair	Red I (6900 A.) – yellow (5890 A.)						Red II (6200 A.) – yellow (5890 A.)					
λ (A.)	6900	6256	6110	6018	5949	5890	6200	6104	6036	5980	5932	5890
P	100	100	100	100	100	100	100	100	100	100	100	100
y_m	2656	2985	3314	3643	3972	4301	3083	3326	3570	3814	4058	4301
Colour pair	Red I (6900 A.) – green I (5300 A.)						Red II (6200 A.) – green I (5300 A.)					
λ (A.)	6900	5996	5797	5627	5450	5300	6200	5926	5756	5598	5439	5300
P	100	98·0	97·0	96·7	96·8	100	100	98·2	96·8	96·4	97·1	100
y_m	2656	3737	4317	5897	6978	8059	3083	4078	5074	6069	7064	8059
Colour pair	Red I (6900 A.) – green II (5100 A.)						Red II (6200 A.) – green II (5100 A.)					
λ (A.)	6900	6010	5767	5471	5187	5100	6200	5930	5715	5430	5177	5100
P	100	85·4	79·7	75·5	83·4	100	100	88·1	80·7	76·5	83·1	100
y_m	2656	3625	4595	5564	6533	7502	3083	3967	4851	5734	6618	7502
Colour pair	Red I (6900 A.) – blue I (4600 A.)						Red II (6200 A.) – blue I (4600 A.)					
λ (A.)	6900	*4969	*5049	*5412	*5645	4600	6200	*4958	*5050	*5460	*5653	4600
P	100	95·0	87·8	75·4	21·8	100	100	70·2	64·7	55·0	31·8	100
y_m	2656	2184	1713	1241	769	297	3083	2526	1969	1412	855	297
Colour pair	Red I (6900 A.) – blue II (4100 A.)						Red II (6200 A.) – blue II (4100 A.)					
λ (A.)	6900	*4970	*5050	*5370	*5613	4100	6200	*4960	*5048	*5420	*5626	4100
P	100	100	100	100	98	100	100	72·9	73·6	72·4	73·0	100
y_m	2656	2135	1613	1091	569	48	3083	2476	1869	1262	655	48
Colour pair	Yellow (5890 A.) – green I (5300 A.)						Yellow (5890 A.) – green II (5100 A.)					
λ (A.)	5890	5766	5647	5525	5403	5300	5890	5749	5571	5323	5162	5100
P	100	98·5	97·5	97·0	97·6	100	100	90·4	85·4	80·0	86·7	100
y_m	4301	5053	5804	6555	7307	8059	4301	5209	6444	7951	8205	7502
Colour pair	Yellow (5890 A.) – blue I (4600 A.)						Yellow (5890 A.) – blue II (4100 A.)					
λ (A.)	5890	6013	*5048	*5598	*5693	4600	5890	6037	*5045	*5548	*5662	4100
P	100	52·7	22·2	17·2	5·7	100	100	52·2	27·0	27·9	28·8	100
y_m	4301	3500	2700	1899	1098	297	4301	3451	2600	1749	898	48
Colour pair	Green I (5300 A.) – blue I (4600 A.)						Green I (5300 A.) – blue II (4100 A.)					
λ (A.)	5300	5172	5037	4922	4806	4600	5300	5178	5037	4911	4765	4100
P	100	84·4	70·0	63·1	57·8	100	100	81·0	67·0	56·2	42·9	100
y_m	8059	6506	4953	3402	1850	297	8059	6457	4854	3252	1650	48
Colour pair	Green II (5100 A.) – blue I (4600 A.)						Green II (5100 A.) – blue II (4100 A.)					
λ (A.)	5100	5043	4980	4909	4814	4600	5100	5043	4977	4900	4781	4100
P	100	94·0	88·0	81·8	74·6	100	100	93·2	85·6	75·8	59·9	100
y_m	7502	6061	4620	3179	1738	297	7502	6394	4807	2950	1130	48
Colour pair	Green I (5300 A.) – green II (5100 A.)						Blue I (4600 A.) – blue II (4100 A.)					
λ (A.)	5300	5247	5197	5158	5126	5100	4600	4560	4510	4450	4340	4100
P	100	97·3	95·8	96·0	97·5	100	100	96	94	91	83	100
y_m	8059	8271	8333	8174	7867	7502	297	239	196	138	82	48

λ = dominant wavelength; P = colorimetric purity %; y_m = relative luminous intensity.

* These mixtures are purples, and the wavelengths marked with asterisks are those of the complementary colours.

Red light is affected progressively less, both in hue and in purity, by addition of blue, green and yellow light. Yellow light is altered in hue most by red or blue light; its purity is reduced considerably by addition of blue light but not appreciably by red. Green light is not greatly altered in hue by addition of up to 30% of light of any colour. Its purity is considerably reduced, more by blue than by red or yellow light. (On mixing green I and green II, the purity drops to about 95% as a minimum.) Blue light is altered in hue most by red, and in purity most by yellow light; the relative intensity is greatly raised by addition of light of any of the other colours. The applicability of these general conclusions to the individual spectra is obvious.

§4. CONCLUSIONS

To be satisfactory, light from coloured flames should fulfil three conditions: (i) it should be of distinctive spectral hue; (ii) it should be saturated; (iii) it should be intense.

Thus the intense emission of the characteristic radiation of some molecular or atomic species which happen to be confined to a narrow range of desired wavelengths would constitute a satisfactory solution. A relevant consideration here is the mode of excitation of the emitting species. On this we have as yet no direct information, but it seems most probable that the excitation in these flames is often purely thermal. If this is so, we can immediately deduce limiting relations for the variation of energy flux with wavelength and with temperature. For, quite generally, the energy radiated in unit time per atom in state n is $A_m^n \boldsymbol{h}\nu_{mn}$, where A_m^n is Einstein's coefficient of spontaneous emission for the change from state n to state m, which results in the appearance of a spectral line frequency ν_{mn}. Now the values of A_m^n for fully allowed transitions are not likely to vary much (cf. Mitchell and Zemansky 1934, Gaydon 1942); for a series, then, of ideal monochromatic emitters of different wavelengths maintained in thermal equilibrium, the intensities will be proportional to $\nu_{mn} \exp(-\boldsymbol{h}\nu/\boldsymbol{k}T)$, where the exponential is simply the Boltzmann factor determining the population of the initial level n of the transition. To obtain relative luminous intensities we multiply this expression by the relative luminosity factor (Judd 1931).

Table 3. Relative Luminous Intensities of Ideal Monochromatic Emitters for Thermal Excitation

Colour	Red	Orange	Yellow	Green	Blue	Violet
λ (A.)	6800	6200	5800	5300	4500	4000
$T=1000°$ K.	$1{\cdot}7\times10^{-7}$	$5{\cdot}2\times10^{-7}$	$2{\cdot}6\times10^{-7}$	$2{\cdot}7\times10^{-8}$	$1{\cdot}1\times10^{-11}$	$2{\cdot}5\times10^{-15}$
$T=2000°$ K.	$6{\cdot}4\times10^{-3}$	$5{\cdot}6\times10^{-2}$	$6{\cdot}2\times10^{-2}$	$2{\cdot}1\times10^{-2}$	$9{\cdot}7\times10^{-5}$	$1{\cdot}6\times10^{-7}$
$T=3000°$ K.	0·22	2·7	3·9	1·9	$2{\cdot}0\times10^{-2}$	$6{\cdot}3\times10^{-5}$

A selection of values is given in Table 3. The big variation with temperature of the intensity at any given wavelength was of course to be expected. More interesting are the facts that ideal red, orange, yellow and green sources have nearly the same relative luminous intensities at constant temperature, and that the blue sources are roughly 100 to 10,000 times less luminous. Any hopes of producing a blue flame depending on thermal excitation comparable in intensity with the other coloured flames are therefore illusory.

Quite generally, however, improvements are to be sought in two directions—in increasing the intensity and in improving the quality of the light. The former implies an increase in the number of emitting species in their upper electronic state, which will generally be brought about by increasing the temperature of the flame. The effect of temperature may, however, be complicated if the emitters are themselves formed by reactions taking place in the flame, e.g. $BaCl_2 \rightarrow BaCl + Cl$; or if decomposition becomes important at high temperature, e.g. $BaCl \rightarrow Ba + Cl$; or if competing equilibria are important, e.g. $BaCl + O \rightarrow BaO + Cl$. For a given emitter, the quality of the colour can be improved by reducing the light-output from undesirable secondary emitters.

Of existing compositions designed to give coloured flames, those giving red flames (based on SrCl and SrO) are the most satisfactory, both from the point of view of intensity and of colour. This is readily understandable in terms of the above discussion. It so happens that only the red systems of SrO and SrCl are of much importance; the blue and near ultraviolet systems of both molecules are developed (as might be expected) only very weakly. Furthermore, the only important secondary emission is that from the resonance lines of potassium which comes from the oxidizing agents generally used—potassium chlorate or perchlorate. These lines lie in the red region of the spectrum and, therefore, contribute to the desired colour. The sodium D lines appear but do not seriously affect the hue or saturation (cf. Table 2). The best results with this type of composition cannot, however, be obtained without purification of the components from barium.

A sodium source is the only likely solution to the problem of finding a yellow flame, and some observations suggest that a similar source may also give a nearly white light of high intensity. It is not difficult to reduce the effects of red and blue light from secondary emitters to negligible proportions.

The distinctive colours which remain are green and blue. Intense flames of these colours do not appear to have been produced. It has been shown above that if the excitation is thermal, the operation of the Boltzmann factor and of the luminosity function combine to reduce the theoretical intensity of a blue flame to quite a small fraction of that expected for, say, an orange flame. Moreover the saturation of a blue flame is much reduced by relatively small proportions of light from secondary emitters; the quality of the colour from existing blue compositions is adversely affected by weaker emission from the A, B and C systems of CuCl, and perhaps by strong emission of the K lines, which, however, have a wavelength near the limit of visibility. Unfortunately it is not easy to find satisfactory alternatives either to the potassium salts or to CuCl as emitter.

The most serious difficulty encountered in attempts to improve the green flames based on BaCl is the emission from the extensive system of BaO. (Small amounts of Na, Ca and Sr also affect the colour adversely, but these could largely be removed by careful purification.) Our results suggest that the ratio BaO/BaCl is greater the more intense the source. Simultaneous improvements in quality of colour and in intensity are not therefore to be expected in sources which rely on emission from BaCl in an atmosphere which contains free oxygen or oxidizing agent. Somewhat surprisingly, there seem to be no very promising alternatives to BaCl among the diatomic molecules, although BO might be useful: it is also possible that a suitable means might be found for exciting the green line of thallium (5350 A.).

ACKNOWLEDGMENTS

This work was carried out for the Ministry of Supply. Our thanks are due to Mr. J. S. Dick for suggesting the investigation and for help in many ways, to Dr. T. Vickerstaff for advice, and to the Chief Scientific Officer, Ministry of Supply, for permission to publish.

REFERENCES

BARROW, R. F., and CRAWFORD, D. V., 1946, *Nature, Lond.*, **157,** 339.
CRAWFORD, D. V., 1944, *B.Sc. Thesis*, University of Oxford.
DAVIS, T. L., 1943, *Chemistry of Powder and Explosives* (New York: John Wiley).
GAYDON, A. G., 1942, *Spectroscopy and Combustion Theory* (London: Chapman and Hall), p. 105.
HAMADA, H., 1933, *Phil. Mag.*, ser. 7, **15,** 574.
HARDY, A. C., 1936, *Handbook of Colorimetry* (Massachusetts Institute of Technology).
JUDD, D. B., 1931, *Bur. Stand. J. Res., Wash.*, **6,** 465.
MITCHELL, A. C. G., and ZEMANSKY, M. W., 1934, *Resonance Radiation and Excited Atoms* (Cambridge: University Press), p. 146.
MOON, P. B., 1936, *Scientific Basis of Illuminating Engineering* (New York and London: McGraw-Hill), p. 483.
PEARSE, R. W. B., and GAYDON, A. G., 1941, *Identification of Molecular Spectra* (London: Chapman and Hall).
WEINGART, G. W., 1943, *Pyrotechnics* (New York: Chemical Publ. Co.).
WIELAND, K., 1945, *Nature, Lond.*, **156,** 504.
WRIGHT, W. D., 1944, *The Measurement of Colour* (London: Adam Hilger), p. 63.

Some Experiments on Photo-ionization in Gases

BY A. A. JAFFE, J. D. CRAGGS AND C. BALAKRISHNAN
Research Department, Metropolitan-Vickers Electrical Co. Ltd.

MS. received 10*th March* 1948, *in amended form* 27*th August* 1948

ABSTRACT. From an investigation of the discharge spread in Geiger counters filled with hydrogen, neon, argon or helium, it is possible to show that in certain circumstances the discharge spread is due almost entirely to photo-ionization in the gaseous counter filling. The results enable the absorption coefficients of the operative radiations to be found.

§ 1. INTRODUCTION

PHOTO-IONIZATION in gases is of interest, particularly in relation to studies of certain forms of electrical discharges in gases (Meek 1940, Loeb and Meek 1941, Weissler 1943). It is also of great importance in upper atmosphere problems (Sayers 1943, Massey 1938, and many papers by Appleton and his collaborators) and finally is relevant to work on Geiger counter discharges (Korff 1946, and many others).

Price (1943) has pointed out the singular lack of experimental data on photo-ionization in the elementary gases, a situation which is in contrast with that obtaining in studies of alkali metal vapours (Ditchburn 1928 etc., Lawrence and collaborators, and others). Page (1939) has published an admirable review (with references) of the theoretical and experimental work up to that time, with particular reference to work on alkalis (Cs, K, etc.), but omitting reference to certain important theoretical papers (Wheeler 1933, Vinti 1933). The later theoretical work is due largely to Massey and his collaborators, particularly Bates (1939, 1947 etc.).

It was thought therefore that further experiments on the elementary gases, employing a variety of widely different techniques, might be of value and because of certain advantages of the method, it was decided first to investigate and refine a Geiger counter method, due originally in a simple form to Greiner (1933) and recently used in experiments on complex gas mixtures (CH_4, C_2H_5OH/argon etc.) by Craggs and Jaffe (1947 a, b), Balakrishnan, Craggs and Jaffe (1948), Alder and his collaborators (1947) and Liebson (1947). The above workers used different techniques as may be seen from reference to the papers.

The method is shown to suffer from certain important limitations which are discussed in § 6, but the present Geiger counter experiments differ from previous comparable work in that photo-emission from the tube walls is eliminated as described below in § 3. The wall effect which was present in Greiner's (1933) and Christoph's (1937) experiments might, it was thought, have exerted an influence and rendered the results spurious.

§ 2. SUMMARY OF PREVIOUS THEORETICAL AND EXPERIMENTAL WORK

Page's review (1939) discusses in some detail the interest evoked in the subject by Kramers' (1923) theory, originally worked out for hydrogenic atoms, and the limited extent to which it agrees with experiments for the alkali metals. Trumpy's later (1931) wave mechanical treatment of the latter was more successful (Page 1939) and gives reasonable agreement with experimental data provided, for example, by Ditchburn and his school (Ditchburn 1928, Braddick and Ditchburn 1934).

Wheeler (1933) and Vinti (1933) provided wave mechanical treatments for helium. Vinti shows, for example, that the atomic cross-section for photo-ionization at the principal series limit ($\lambda = 504$ A.) is about $8{\cdot}4 \times 10^{-18}$ cm², falling off to about one-tenth of this value at about 180 A. The cross-section for the carbon Kα radiation (44·6 A.) was measured by Dershem and Schein (1931) and agrees satisfactorily with Vinti's calculated value.

The atomic cross-section σ_a is related simply to the linear absorption coefficient μ (pressure p) by $\sigma_a = \mu/N_0$, where N_0 is the number of atoms per cm³ (pressure p), and μ is given by $I = I_0 e^{-\mu x}$, where I_0 and I are the incident and emergent intensities for a thickness x of absorber.

Theoretical treatments for atomic hydrogen are due to Suguira (1927), Gaunt (1929), Menzel and Pekeris (1935) and others. Massey (1938) shows $\sigma_a \simeq 0{\cdot}6 \times 10^{-17}$ cm² at the series limit.

Since the writing of Page's review, Bates' paper (1939) appeared in which results of great interest are described. The continuous absorption coefficients per $2p$ electron for B, C, N, O, F, and Ne are calculated for various ejected electron energies. The introduction of the appropriate weighting factors gives the following values of absorption cross-sections at spectral heads. (Bates, private communication).

Table 1

Atom	B	C	N	O	F	Ne
Cross-section (10^{-17} cm²)	1·9	2·4	2·2	0·45	0·61	0·58

It will be seen that Ne is the only convenient element with which the above results may be checked. Bates' results show that the Ne absorption varies much less with energy than does the helium absorption worked out by Wheeler (1933) and Vinti (1933). Experiments to check these He and Ne results are in progress at the University of Liverpool, using a different method from that to be described below.

Experimental data on the elementary gases are scarce. Greiner (1933) and later Christoph (1937) used a Geiger counter method in which the discharge of one counter (a) was allowed to initiate a discharge in another counter (b) enclosed in the same chamber, by virtue of photo-ionization produced in (b) by energetic photons from the discharge in (a). Their results are summarized in Table 6. Raether (1938) described a novel method in which he measured photo-ionization caused in the gaseous filling of a Wilson cloud chamber, by photons passing through a very thin cellophane window from an external corona discharge into the chamber. Geballe (1944) measured the absorption coefficient for highly absorbable radiations produced in a Townsend discharge and detected photo-electrically at brass surfaces placed at different distances from the discharge. These results are also given in Table 6 and the discussion in § 6. Finally, Schneider (1940) measured absorption coefficients in air for wavelengths lying between 380 and 1600 A., using a vacuum spectrometer.

§ 3. APPARATUS

A new type of divided cathode, single anode counter—with provision for collimating the photons originating in the discharge—was constructed (Craggs and Jaffe 1947 a, b). It consists of six copper cathodes, each 5 cm. long, 2 cm. internal diameter and spaced 1·2 cm. apart, with an anode wire of 5 mil tungsten (Figure 1).

In order to eliminate the cathode effect by collimating the photon beam to the neighbourhood of the wire, a window of C40 glass (which seals well to Fe/Ni/Co alloy), was attached to the front of each except the first cathode. The central hole in each window was 3 mm. in diameter in the beginning. Later, that for cylinder B was replaced with another window with a hole of diameter 1 mm. only. The alignment of the five windows was carried out with great care and the collimation checked before the wire was inserted and the tube sealed at the ends.

A special counter was used in order to check that no radiation passed through the glass of the windows used in the previous counter. This consists of two separate compartments separated by a circular piece of C40 glass of the same thickness as that used in the six-cylinder window counter. A cylindrical piece of Kovar wire is fused to the glass through a hole in the centre. The copper cathodes (above dimensions) are spaced 10 cm. apart.

The quenching circuit used, shown in Figure 2, is similar to that of Neher and Harper (Craggs and Jaffe 1947 b). Careful tests were made to ensure that the coincidence circuits were symmetrical and that both channels were identical in sensitivity. The two-channel quenching circuit, which was arranged to take pulses from the cathodes of the divided counter, supplied the pulses to a scaling circuit and mechanical counter. Gases of cylinder grade were used.*

* The experiments were intended to give information relevant to the conditions of sparks in cylinder gases, but it is necessary to stress that the observed μ values (Table 6) could be given by absorption in impurities (having $\mu' \sim 40$ cm^{-1} at 10 cm. Hg) being present to the extent of $\sim 0{\cdot}25\%$. The H_2 and Ne results should be affected less than those with A since in the former cases the main impurities have higher ionization potentials than the main gas (see also Raether 1938).

§4. PRELIMINARY TESTS

It was necessary to determine the resolving time of the coincidence circuit used, in order to check that it was short enough to reduce random coincidences to a reasonably low level, but was not comparable with or less than, the time for discharge spread, which would be of the order of 10^{-8} sec. for photon induced spread. Two similar separate counters, each equivalent to one section of the counter of Figure 1, were set up 10 cm. apart. By taking the individual counts N_1 and N_2, when one of them is irradiated with a collimated source of gamma rays from radium and the chance coincidence counts C in the same conditions, the resolving time, τ, can be calculated from the equation $C=2N_1N_2\tau+K$, where K, the true cosmic ray coincidence count, may be eliminated by using two sets of data. N_1 and N_2 included the counter background in each case. By reducing the operating voltage of one counter just below its starting potential, the absence of circuit pick-up effects was checked.

The following results were obtained in a typical experiment with two separate counters:

(*a*) With no gamma radiation: $N_1=648/\text{min.}$, $N_2=549/\text{min.}$, $C=2{\cdot}8/\text{min.}$

(*b*) With gamma radiation: $N_1=906/\text{min.}$, $N_2=1321/\text{min.}$, $C=6{\cdot}9/\text{min.}$

τ was calculated and found to be about $1{\cdot}5\times10^{-4}$ sec.

The time constant involved in the circuit of Figure 2 is such that coincidences are missed at high counting rates, and the effect was shown by calibration curves relating to the number of counts per cylinder and the coincidence counts. To avoid distortion of the discharge spread curves (Figures 4–6) by such a spurious effect, the runs for the latter figures were all made at as constant a counting rate as possible.

Experiments were conducted with the test counter to check that the glass used for the windows is the six-cylinder counter was opaque to the radiation originating in the gas discharges. The two-chamber counter was filled with argon and hydrogen in turn at various pressures and one of the cylinders was irradiated with a collimated source of gamma rays. Determinations were made of the counting rates of this irradiated cylinder and of the other in the neighbouring compartment together with the coincidence counts for the two cylinders. These rates, when compared with the corresponding rates in the case of two individual counters arranged at the same distance apart, show that the coincidence rate for the two-chamber counter is equal to the chance coincidence for the two separate counters. Hence it may be safely assumed that the glass used for the windows in the six-cylinder counter (Figure 1) is opaque to the radiation originating in the discharge and so the photon beam is not irradiating the cathodes. Table 2 gives a summary of the data observed. The counting rates for two cylinders (in the six-cylinder counter) whose distance of separation is approximately equal to that of the cylinders in the two-chamber counters are also recorded in this table. This shows the amount of discharge spread at the distance, i.e. at approximately 10 cm.

Previous work (Craggs and Jaffe 1947 b and references cited there to earlier experiments by other workers) showed that the discharge spread in H_2 and A for divided cathode counters of the dimensions shown in Figure 1, but without collimating windows, was approximately 100%. Greiner (1933) who did not employ collimation found less than 100% spreading, but this discrepancy is explicable by different cathode work functions in Greiner's and the present experiments.

§ 5. EXPERIMENTS WITH SIX-CYLINDER WINDOW COUNTER

The experimental arrangement is shown diagrammatically in Figure 3. The counter was filled, at a definite pressure, with the gas under investigation and mounted before a lead collimator so that cylinder A was irradiated with a collimated beam of gamma rays from 0·5 mgm. of radium embedded in the lead collimator. Cylinders A and B were connected to the counting channels and their counting

Table 2

Type of counter	Gas filling	Pressure (cm. Hg)	(a)	(b)	(c)
Two separate	Argon	10	1321	906	6·9
counters	Argon	5	1681	1496	21
Two-chamber	Argon	10	638	594	5
counter	Hydrogen	10	640	415	8
	Hydrogen	5	1209	80	3
Six-cylinder	Argon	20	419	208	87
window counter		10	1015	1071	661
(Figure 1)		5	807	754	607
		2·5	600	494	384
	Hydrogen	20	443	152	135
		10	466	323	307
		5	382	350	332
		2·5	361	360	331

(a) Counting rate/min. for cylinder irradiated with gamma rays.
(b) Counting rate/min. for other cylinder.
(c) Coincidence counting rate/min.

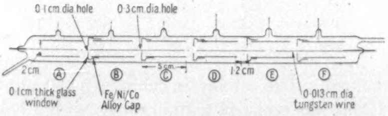


Figure 1. 6 cylinder window counter.

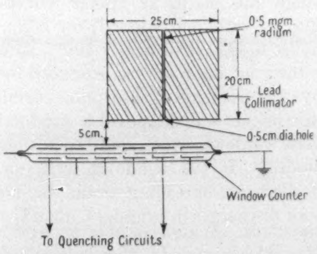


Figure 3. Diagram of experimental arrangement.

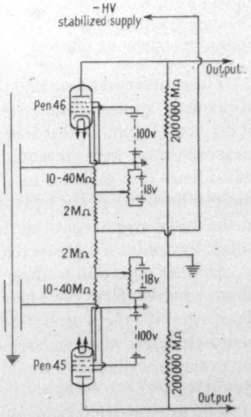


Figure 2. Two-channel external quenching circuit.

rates were taken. The coincidence rate for the two cylinders was also recorded. The measurements were repeated with cylinder A and each of the successive cylinders, C, D, E and F connected to the set.

If N_1 and N_2 are the counting rates of two cylinders and N_C the coincidence counting rate, the total of the true counts of the two cylinders will be $N_1+N_2-N_C$, since the spreading of a pulse from the first to the second will cause an apparent increase in the counting rate of the second and so also for the pulses which spread from the second to the first. The percentage pulses which cause coincidence is $100N_C/(N_1+N_2-N_C)$.

The percentage spreading was calculated for each cylinder and plotted against the distance of that cylinder from the irradiated cylinder.

In order to check the absence of electron diffusion and to ascertain if any spurious charges which might affect the results were accumulated on the intermediate cylinders when the remote ones were counting, counting rates were taken with the intermediate cylinders (*a*) floating, (*b*) at a potential less than the starting potential and (*c*) shorted to the wire. The data are recorded in Table 3 and it is seen that the results are not affected in any way by leaving the intermediate cylinders floating.*

The curves obtained were flat topped, showing that there was 100% spreading for the first two or three cylinders. This may be due to an excess of photons generated in cylinder A, since one photo-ionizing photon will give a count.

Table 3

Intermediate cylinders	Floating	At 700 v.	Floating	At 500 v.	Floating	Shorted to wire
Counting rate/min. of A	948	947	351	374	170	183
Counting rate/min. of B	942	958	358	352	189	167
Coincidence counting rate/min.	587	624	151	147	116	123

The counter was now slightly modified. The window in cylinder B had a hole of diameter approximately 3 mm. This was replaced with another, with a hole of diameter 1 mm., so that less photons might spread into the other cylinders, and measurements were repeated. Although the discharge spread curves (i.e. coincidence rate–cylinder number (Figure 1) with A irradiated) for argon were straight lines, those for hydrogen were flat topped with nearly 100% spreading in the initial stages, showing that more than one ion-pair was produced for each pulse in cylinder A and so excluding any measurements of absorption coefficients.

Another objection to these early experiments was the possible production of photo-emission from the edges of the glass windows, which could then trigger off discharges in the corresponding cylinders. If the windows were emitting photo-electrons, then an asymmetry in discharge spreading would be noticed, for example, cylinder A would give more discharges in cylinder C than C would give in A.

The asymmetry in spreading was detected experimentally. Cylinder A was irradiated with the collimated source of gamma rays, and the counting rate, when this cylinder alone was connected to the set, was taken. Next cylinder E was also connected to the set, and the counting rate for A with E and E with A and the coincidence rate for them were found. The collimator was then moved so that the gamma ray beam irradiated cylinder E and the counting rate of E (with A disconnected) was adjusted to be the same as that of A (with E disconnected), by manipulating the position of the radium in the collimator. Then the counting rate of E alone, of E with A connected to the set and the coincidence counting rate were measured. Since the individual counting rates of A and E were adjusted to be equal, if the spreading was symmetrical, the counting rate for A with E and E with A as well as the coincidence rates in the two cases of irradiation should be equal.

---

* Later experiments on the velocity of discharge propagation (to be published) indicate that other diffusion phenomena are not important. One of us (J. D. C.) wishes to acknowledge a valuable private communication from Professor L. B. Loeb on these matters.

Table 4 shows the data and it is easily seen that the spreading is asymmetrical. The degrees of collimation for the two directions of discharge spread should be sensibly equal.

Table 4. Counting Rates for various Conditions

	A, with E disconnected	E, with A disconnected	A, with E connected	E, with A connected	Coincidence counts
A	426	41	423	330	247
E	33	453	87	513	69

First column indicates cylinder irradiated.
Figures give number of counts/minute.

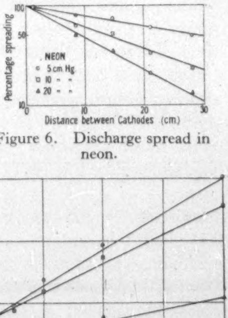


Figure 6. Discharge spread in neon.

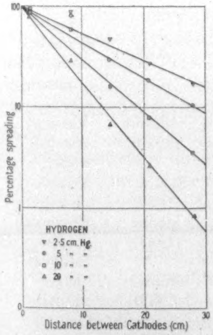


Figure 4. Discharge spread in hydrogen.

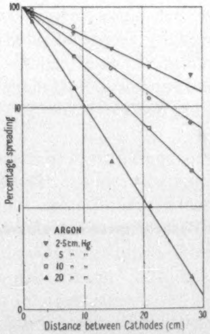


Figure 5. Discharge spread in argon.

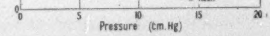


Figure 7. Absorption coefficients (arbitrary units) plotted as a function of pressure.

In view of this asymmetry in spreading all subsequent experiments were modified to eliminate this window-edge effect. Instead of irradiating cylinder A and measuring the spreading in the other cylinders, each of the cylinders B, C, . . . F were irradiated successively with the collimated beam of gamma rays, and the spreading to cylinder A was measured in each case. The spreading curves for hydrogen, argon and neon are plotted in Figures 4, 5 and 6 respectively, at various pressures; Figure 7 shows a plot of the slope of the spreading curves against pressure. The apparent absorption at zero pressure is presumably due to the imperfect collimation. Table 5 gives a summary of the starting voltage and the working voltage in each case. Experiments made in the initial stages show that the absorption does not vary with overvoltage although there is greater spreading at higher voltages.

Table 5

Gas	p	v_S	v_W	Gas	p	v_S	v_W
Argon 98 %	2·5	580	620	Hydrogen 99 %	2·5	720	760
(∼2 % N_2)	5	760	800	(∼1 % O.)	5	880	920
	10	920	960		10	1160	1200
	20	1220	1260		20	1600	1640
Neon 98 %	5	380	420	Helium 98 %	10	1100	1200
(∼2 % He)	10	400	440	(∼2 % Ne)	3	830	930
	20	440	480				

p=pressure (cm. Hg); v_S=starting voltage; v_W=working voltage.

The values of absorption coefficient μ (see § 2 for definition) may be derived from Figures 4, 5 and 6 and are given (for 10 cm. Hg pressure) in Table 6; they were quite close to those obtained in the presence of emission from the edges of the glass collimating windows. Data for helium were also obtained in some preliminary experiments, but owing to lack of helium, further data without the window effect

Table 6. Values of Absorption Coefficient μ for $p = 10$ cm. Hg

H_2	Air	O_2	H	N	O	He	Ne	A	p	Author	Experimental method and remarks
—	∼50	—	—	—	—	—	—	—	∼0·1	Schneider (1940)	Vacuum spectrometer
0·11	0·26	0·66	—	—	—	—	—	—	14–50	Raether (1939)	Wilson chamber
55	—	—	—	—	—	—	—	—	0·1	Geballe (1943)	Photo-cell (brass cathode)
0·18	—	—	—	—	—	—	—	—	∼10	Greiner (1933)	Geiger counter
3·6	7·0	—	—	—	—	—	—	—	∼1·5	Christoph (1937)	Geiger counter
—	—	—	22	—	—	—	—	—	—	Massey (1938)	Calculated at series limit
—	—	—	—	80	16	—	21	—	—	Bates (1939)	Calculated at series limit
—	—	—	—	—	—	31	—	—	—	Vinti (1933)	Calculated at series limit
0·12	—	—	—	—	—	—	0·051	0·13	2·5–10	Present authors	Geiger counter

p=gas pressure (cm. Hg) used in experiments.

(see above) were not obtained. However, it seems reasonable to state that the absorption coefficients for helium lie between those of hydrogen and neon.

In some experiments with a long counter (Figure 8) cylinder A was irradiated with the collimated radium source and the discharge spread in cylinders D, E, . . . H was measured in the usual way. Owing to the long distances involved in this counter, there was perceptible spread only in neon and hydrogen at low pressures. At higher pressures in these gases as well as in argon there was almost complete absorption and practically no spreading at such great distances.

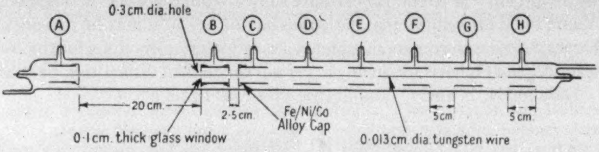


Figure 8. Long collimating counter.

§ 6. DISCUSSION OF RESULTS

Previous experimental workers have often tended to ignore the theoretical work on He (Vinti 1933) and atomic H (Table 6 and elsewhere) and arranged experiments which obviously would not have permitted such radiations to be measured, due to the use of excessive path lengths, etc. For example, taking μ_{air} as 50 cm^{-1} (Schneider 1940, Table 6), $I_0/I \simeq 100$ for $x = 0{\cdot}1$ mm. at 1 atmosphere pressure and the absorption at 1 cm. Hg and $x \simeq 1$ cm. would also be about 100 : 1, so that counter experiments could not be expected to yield much useful information on such absorbable radiations. The same conclusion could have been reached from the 1933 work on He and from Bates' (1939) paper. These high values of absorption coefficient lend qualitative but strong support to the streamer theory of spark breakdown proposed by Loeb and Meek (1941 and references there cited).

However, since the values of σ_a with frequency for He and Ne are different, it was hoped that Geiger counter experiments might demonstrate such differences, and indeed for this reason it was also thought that the neon values for σ_a might be useful because, since a gaseous discharge (Geiger counter or Townsend [Geballe] type) would generate photons with a finite width of spectrum, the effect of such a spectrum in determining the "mean" absorption coefficient of the radiations should be less in Ne than in other gases.

If a suitable monochromatic source of photo-ionizing radiation could be found, and the x-rays used by Dershem and Schein were of too short a wavelength to give much valuable information on absorption near the series limit, then the filtering effect of distance would clearly be of less importance.

It is concluded that the vacuum spectrometer method has exceptional value (Table 6) since low gas pressure ($<$1 mm. Hg) may be used for absorption, and most electrical discharge devices (e.g. Geiger counter) need a higher gas pressure than this, although Geballe's Townsend discharge could be operated at approximately 1 mm. Hg. Geballe's work might be criticized that the emissions from two (different) photo-electric surfaces of brass (not degassed) at different distances from the photon generating discharge were compared without direct comparison of the work functions of the two surfaces.

§7. CONCLUSIONS

In conclusion, therefore, it appears that the present (Geiger counter) method and those of Raether, Greiner etc. cannot be expected to give true values of the absorption coefficient for photo-ionizing radiations, due to excessive filtration effects. The existence, in such discharge devices as Geiger counters, of radiations having $\mu \sim 0{\cdot}1\ \text{cm}^{-1}$ at 10 cm. Hg pressure shows (from the work of Wheeler (1933) and Vinti (1933) for helium) that radiations having $\lambda \sim 50$ A. may be present in such discharges.* It is unfortunate that photo-ionization cross-sections for the common gases H_2, N_2 and O_2 have not, owing to certain theoretical difficulties, been worked out. Data for H, N and O (atomic) are given in Table 6.

ACKNOWLEDGMENTS

The authors are indebted to Professor H. S. W. Massey and to Mr. D. R. Bates for much advice on theoretical matters; Mr. Bates brought several valuable references to their notice. They also wish to thank Mr. A. J. Knowles for the loan of counting equipment, Mr. F. R. Perry for his continued support of the work, and Sir Arthur P. M. Fleming, Director of Research and Education, and Mr. B. G. Churcher, Manager of the Research Department, Metropolitan-Vickers Electrical Co. Ltd., for permission to publish this paper.

REFERENCES

Alder, F., Baldinger, E., Huber, P., and Metzger, F., 1947, *Helv. Phys. Acta*, **10,** 73.
Balakrishnan, C., Craggs, J. D., and Jaffe, A. A., 1948, *Phys. Rev.* **74,** 410.
Bates, D. R., 1939, *Mon. Not. R. Astr. Soc.*, **100,** 25; 1947, *Proc. Roy. Soc.* A, **188,** 350.
Braddick, H. J. J., and Ditchburn, R. W., 1934, *Proc. Roy. Soc.* A, **143,** 472.
Christoph, W., 1937, *Ann. Phys. Lpz.*, **30,** 446.
Craggs, J. D., and Jaffe, A. A., 1947a, *Nature, Lond.*, **159,** 369; 1947b, *Phys. Rev.*, **72,** 784.
Dershem, E., and Schein, M., 1931, *Phys. Rev.*, **37,** 1238.
Ditchburn, R. W., 1928, *Proc. Roy. Soc.* A, **117,** 486.
Gaunt, J. A., 1929, *Philos. Trans.* A, **229,** 163.
Geballe, R., 1944, *Phys. Rev.*, **66,** 316.
Greiner, H., 1933, *Z. Phys.*, **81,** 543.
Korff, S. A., 1946, *Electron and Nuclear Counters* (New York: Van Nostrand).
Kramers, H. A., 1923, *Phil. Mag.*, **46,** 837.
Liebson, S. H., 1947, *Phys. Rev.*, **72,** 602.
Loeb, L. B., and Meek, J. M., 1941, *The Mechanism of the Electric Spark* (Stanford University Press, U.S.A.).
Massey, H. S. W., 1938, *Negative Ions* (Cambridge: University Press).
Meek, J. M., 1940, *Phys. Rev.*, **57,** 772.
Menzel, D. H., and Pekeris, C. L., 1935, *Mon. Not. R. Astr. Soc.*, **96,** 77.
Page, T. L., 1939, *Mon. Not. R. Astr. Soc.*, **99,** 385.
Price, W. C., 1943, *Rep. Prog. Phys.*, **9,** 10.
Raether, H., 1938, *Z. Phys.*, **110,** 611.
Sayers, J., 1943, *Rep. Prog. Phys.*, **9,** 52.
Schneider, E. G., 1940, *J. Opt. Soc. Amer.*, **30,** 128.
Suguira, Y., 1927, *J. Phys. et Radium*, **8,** 113.
Trumpy, V., 1931, *Z. Phys.*, **71,** 720.
Vinti, J. P., 1933, *Phys. Rev.*, **44,** 524.
Weissler, G. L., 1943, *Phys. Rev.*, **63,** 93.
Wheeler, J. A., 1933, *Phys. Rev.*, **43,** 258.

---

* Such radiations could be produced even in pure gases by radiative recombination or by electron excitation in certain cases of multiply ionized atoms.

The Excitation and Transport of Metal Vapour in Short Sparks in Air

By G. C. WILLIAMS, J. D. CRAGGS and W. HOPWOOD
Research Department, Metropolitan-Vickers Electrical Company Limited, Manchester

MS. received 31*st October* 1947, *and in amended form* 21*st April* 1948

ABSTRACT. The paper describes a study of the excitation temperature in certain spark discharges of accurately known current characteristics. The excitation temperatures are found by measuring intensity ratios for certain spectral lines where the relevant transition probabilities are known, and are comparable with earlier work by Ornstein and his collaborators on arc discharges.

The excitation temperatures are, as would be expected, higher than those generally found for arc conditions.

Certain peculiarities relating to the evaporation of electrode metal are described. In particular the presence of discontinuous evaporation was noticed, and it was found that different electrode metals showed considerable variations in their behaviour in this respect.

§ 1. INTRODUCTION

The study of phenomena occurring in spark gaps is of interest for several reasons; there are many technical applications of spark discharges for various and widely differing purposes, such as the ignition of explosive gaseous mixtures in internal combustion engines and the spectroscopic analysis of many substances. The behaviour and function of the metallic electrodes used in these circumstances are clearly of considerable importance, and many aspects of the discharge-gas–electrode-metal interaction are little understood.

The present experiments were undertaken in order to obtain preliminary information on excitation conditions in spark discharges in air whose electrical characteristics were defined more clearly than in previous comparable work (see §§ 2 and 6); in particular, to give values of the excitation temperature in the discharges, using spectral lines of Ba and Mg for which the transition probabilities are known from separate experiments with arc discharges. The latter work is due largely to Ornstein and his collaborators.

Extensions of this work in various directions, particularly to the case of more arc-like discharges, are taking place.

§ 2. SUMMARY OF PREVIOUS EXPERIMENTS

Unlike the excitation conditions in pure arcs (e.g. Ornstein and Brinkman 1934) and even in glow discharges, the modes of electron–ion or electron–atom interactions in pure spark discharges have been little studied and are not so well understood, probably because of the relatively greater complexity of the discharge. It is difficult, in the first place, to define a pure spark but for present purposes the expression is taken to mean a transitory gaseous discharge in which (*a*) the voltage gradient is high and is changing with time, (*b*) *thermal* equilibrium is approached to an unknown, but probably limited, extent, (*c*) the general electrical state can be described as being between that of some forms of corona (streamers etc. (Loeb and Meek 1940)) and glows or arcs. The general physical characteristics of sparks are well known, viz. brightness, tortuous nature etc.

Previous work on such discharges, with particular relevance to the study of spectroscopic analysis, is due to Kaiser and Wallraff (1939), Langstroth and McRae (1938 a, b) and many others. A useful and accurate review of the subject is by Llewellyn Jones (1945). In the above researches, oscillatory discharges have been used and photographic records of spectra generally give an integrated record of the emission for the whole wave train. This may be misleading, since the spark discharge probably only obtains for the first half-oscillation and tends to become an arc-like discharge during the latter oscillations of the train. For example, Langstroth and McRae (1938 a) used a circuit which gave a damped train of waves with a period of 100 μsec., and found from observations of the well-known cyanogen bands, and from certain tin lines (2850, 3009, 3262, 3175 A.) that thermal equilibrium was established during the discharge time, i.e. the excitation was thermal with approximately equal gas, electron and ion temperatures. The temperature found was about 9,500° K. However, it was pointed out (Langstroth and McRae 1938 b) that the potential difference across the spark gap, for every half-oscillation after the first, was about 300 volts falling rapidly to about 35 volts in a fraction of a period. Hence it is not surprising, in the present authors' opinion, that thermal equilibrium was established.

A few measurements (Langstroth and McRae 1938 a) were made with the same circuit modified to give a damped wave train with a period of about 1 μsec. It was found that thermal equilibrium was not established since temperatures deduced from the tin lines lay between 8,700° and 11,000° K. Apparently no measurements with the cyanogen bands were made. However, from these measurements the departure from thermal equilibrium (see § 6 for a discussion of energy distributions and statistical equilibria in certain high current density discharges) is apparently not very large, considering the difficulty of interpreting the data and the uncertain knowledge of the exact form of the discharge. This may again be due to the nature of the discharge, and the finite length of the oscillatory wave train.

Recurrent sparks were used carrying square waves of current (Craggs, Haine and Meek 1947, Craggs and Meek 1946). In this case the wave form of the spark current is accurately known. About 20–30 kv. (accurately measurable) is applied to a spark gap arranged in series with a special 3-electrode trigger gap (Craggs, Haine and Meek 1947) so that the test gap breaks down in much less than 1 μsec. The following current wave is sensibly rectangular (see Figure 4 (*e*)) and of controllable duration, e.g. 1–9 μsec. in the present experiments (§ 3). The recurrence rate is controllable and is usually about 50/sec. The voltage across the gap falls from its initial value (say 20 kv.) to about 250 volts (in air) in about 0·5 μsec. (Craggs, Haine and Meek 1947, Figure 27). The sparks were observed with electron multiplier tubes, coupled to a cathode-ray oscillograph, and with a rotating mirror camera.

Further work will be carried out with even shorter discharge times and a separate detailed investigation of the voltage characteristics of short sparks is in progress.

§ 3. DESCRIPTION OF THE APPARATUS

The square current pulses were produced by the charging and discharging of an artificial transmission line using a circuit described by Craggs, Haine and Meek (1947). The oscillograph and spark firing circuits were electrically synchronized.

The optical system for spark observation was arranged so that light from the spark was focused on the collimator lens. The telescope eyepiece was replaced by another slit in the focal plane of the spectral lines; the electron multiplier, which was coupled through an amplifier to the cathode-ray oscillograph, was mounted just beyond the telescope slit, and the telescope slit width was adjusted to give adequate resolution. Further details of the amplifier are given by Craggs and Hopwood (1948).

The rotating mirror used for observation of the vapour clouds emitted by the spark electrodes was such that a resolution in time of 40 μsec/cm. was obtained on the photographic plate. The plates were developed and microphotometered with care using normal techniques, such as continuously brushing with cotton wool during development.

§ 4. EXPERIMENTAL RESULTS

In order to determine excitation temperatures by the measurement of spectral lines the transition probabilities of the latter must be known. Ornstein and Brinkman (1934) and others have measured the transition probabilities for many lines produced in arcs where, as can be shown by observation of the CN bands, thermal equilibrium is obtained and the temperature may be found. The excitation temperature is found from the CN bands or in other ways, and this knowledge enables the relative transition probabilities of lines, whose relative intensities are known, to be found. The operative relation is

$$I_1/I_2=(A_1G_1\nu_1/A_2G_2\nu_2)\exp\{(E_2-E_1)/\boldsymbol{k}T\}, \qquad \ldots\ldots(1)$$

I_1 and I_2 are the measured intensities of two lines of frequencies ν_1 and ν_2 and relative transition probabilities A_1 and A_2, G_1 and G_2 are the respective statistical weights, E_1 and E_2 are the energy values (ergs) of the initial levels for the two transitions, and $\boldsymbol{k}$ and T are Boltzmann's constant and the gas temperature respectively. We shall show later that even when thermal equilibrium is not attained, T may be almost equal to T_e (electron temperature) if certain conditions are satisfied by the discharge.

The main purpose of the investigation was to derive information on spark channels in air between metallic electrodes, using the known transition probabilities, found from the work on the arc discharge. The present study is limited to visible radiation; Ba and Mg seemed to be the most suitable metals of those whose transition probabilities have been measured. The work on Ba and Mg was carried out by Mason (1938), Kruithof (1943) and Kersten and Ornstein (1941). Tests on other metals, e.g. Ca and Cd, whose transition probabilities were known, showed that they were not so suitable for the present preliminary study.

The electron multiplier experiments are first described. The spectral response (Figure 1) of the multiplier was such that many red lines (e.g. some of those of BaI) could not be measured. The Mg triplets 3838, 3832 and 3829 A. and 5184, 5173 and 5167 A. were measured as groups of three added lines since the available dispersion was too small to give adequate resolution into the separate component lines. Figure 2 shows some typical results with these Mg triplets (approximately 100 amp. pulse for 4·5 μsec. with a 3-mm. gap in air between pointed electrodes).

When tests with Ba metal electrodes were made, the spark-to-spark fluctuations in light output, for constant current pulse shape, were too great to allow satisfactory oscillograms to be taken. To find the source of these fluctuations, the spark gap was moved across the optical axis of the system and the peak heights of the light-emission–time oscillograms (e.g. Figure 4 (*d*)) noted for each of a number of fixed spark conditions. From the plot of peak light intensity against displacement from the axis it was possible to show that the effect of random spark movement (with fixed electrodes on the optical axis) on the oscillogram variations was small. The observed spark light fluctuations corresponded to off-axis movements of about 8 mm. in a typical case (Figure 4 (*d*)) whilst, in practice, the maximum random spark movements were only about 2 mm. from the mean (axial) position. Much greater fluctuations than those of Figure 4 (*d*) were often observed. It seems therefore that the great observed fluctuations in the amount of light (at any particular wavelength) emitted in a single spark of the present character must be due to electrode effects. Ba serves as an extreme case, where the fluctuations are most pronounced. It must be emphasized again that the current wave from spark to spark showed no fluctuations.

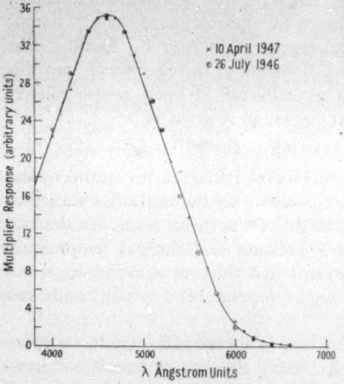


Figure 1. Spectral response of the electron multiplier.

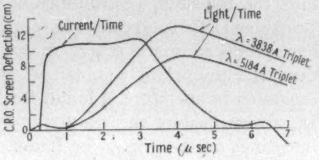


Figure 2. Light/time and current/time oscillograms taken with Mg electrodes and 4 μsec. current pulse.

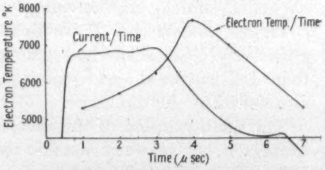
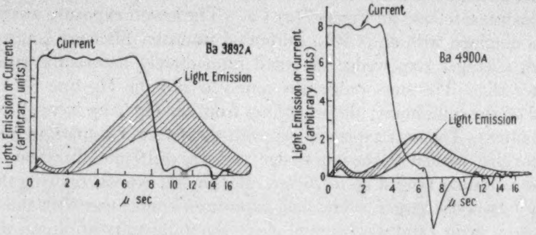


Figure 5. Electron temperature/time for a spark with Mg electrodes. The electron temperatures are derived from the ratio of intensities of the MgI triplets. $\lambda = 3838$ A. and $\lambda = 5184$ A.

It is necessary to show that the light emission fluctuations as in Figure 4 (*d*) are not due to multiplier noise. The relative noise fluctuations (R_t) are given by (Dieke *et al.* 1946) $R_t = 1/\sqrt{(nt)}$, where t is the time interval considered, and n is the number of electrons flowing per second. The multiplier photo-cathode current is not less than 20×10^{-12} amp. and for $t = 10^{-6}$ sec., $R_t \simeq 9\%$. The fluctuations shown in Figure 4 (*d*) ($\sim 30\%$) are much greater than this and are in fact themselves much less than the fluctuations observed in other cases. Further confirmation of the relative unimportance of multiplier noise is provided by observations on hydrogen sparks which show fluctuations considerably less than 9% in the conditions of the present experiments.

The use of various barium salts in drilled electrodes gave no improvement, but specially prepared alloys Ba/Cd (20%/80%) and Ba/Mg (20%/80%), gave successful results. The light–time oscillograms still showed fairly large fluctuations, but enabled satisfactory readings to be taken. Fitzgerald and Sawyer (1934) used a Ba/Al alloy.

Tracings of oscillograms for certain Ba lines, for two different pulse lengths are given in Figures 3 (*a*), (*b*). Figure 4 (*a*)–(*f*) show reproductions to 3/5 full size of oscillograms for CdI (4678, 4799 and 5085 A.), of BaII (4934 A.), and of the current pulse and timing oscillation (2·8 μsec. periodic time).



(*a*) (*b*)

Figure 3. Oscillograms, light emission/time and current/time for sparks in air between Ba electrodes.

The light emitted by the sparks was shown to increase with time during the discharge and to reach its maximum intensity about 1 μsec. after the current had begun to fall (Figure 2). The light is still intense when the current has fallen to zero, and diminishes gradually thereafter.

The excitation temperature derived from the relative intensities of the above mentioned Mg triplets is plotted as a function of time in Figure 5. The results are discussed in § 6 (from which, for these discharges, electron temperatures $\equiv$ excitation temperatures).

It was found that ionized lines (BaII) reached their maximum strength of emission some 2–3 μsec. later than neutral lines, from whatever element (Ba, Cd, Mg) the latter were produced. The small peaks shown at the beginning of these and other oscillograms (Figures 4 and 6) are due to electrical pick-up. The effect was easily shown (since it arose from the current used to fire the 3-electrode triggering gap placed in series with the test gap) not to effect the oscillograms after the first microsecond. The spark currents are so high and the current pulses so

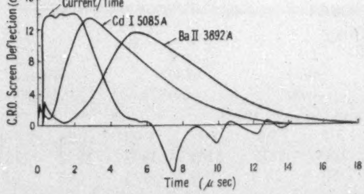


Figure 6. Comparison of light/time oscillograms for neutral and singly ionized lines.

sharp and short in time that slight pick-up on the amplifiers proved very difficult to eliminate, since the overall amplification of the muliplier and amplifier system was 10^8–10^9.

Spectra of Ba, Ca and Mg obtained with the short sparks are given in Figure 7 (*b*), (*c*), (*d*) with Hg comparison spectra. The lines were identified from those of Hg by the use of standard interpolation methods. The procedure for comparing line intensities was as follows for Figure 7 (*a*), Ca. The lowest exposure was taken first, and was obtained with an N.P.L. calibrated tungsten filament lamp (2700° K.) and with a Hilger step wedge mounted immediately before the spectrometer collimator slit. The step wedge was removed and the Hg line 5461 A. superimposed on the continuum; the other lines from the Hg lamp were absorbed by a suitable filter. The next exposure was with the 4 μsec. Ca spark in air (35 μsec. collimator slit), again with the step wedge, and was of 10 min. duration. In other cases the exposures ranged up to 70 min., the longest exposures being those taken with the 1 μsec. discharges. The final exposure (1 min.) was with the Ca spark, without step wedge, and was taken to check the uniformity of illumination along the length of the slit.

The distance of each microphotometered line was measured relative to the superimposed Hg line and corresponding points were then taken along the continuum for density measurements. From the data thus obtained, using the six steps of the step filter for the metal lines and the calibration continuum, the spectral response of each plate and thus the relative line intensities were determined. The microphotometer technique followed normal practice.

Since it appears (Hopwood and Craggs 1947) that the spectral response of the plates used in these experiments is different for the spark light ($\simeq$1–10 μsec. in duration), and for steady light, comparisons were only made between lines whose wavelengths differed by amounts ($\leqslant$400 A.) too small to cause appreciable errors. The existence of this phenomenon (intermittency effect on spectral response) is usually and justifiably ignored since it does not occur in most spectroscopic work.

Table 1

BaI (Kruithof 1943)

Initial level (ev.)	3·80	3·79	3·16	3·16	3·15	3·14
Wavelength (A.)	5778	5519	6063	5997	6019	6111
Ag (relative)	6·55	3·41	1·95	0·84	1·00	3·06

BaII (Mason 1938)

Initial level (ev.)	2·5	5·22	5·22	5·66
Wavelength (A.)	4934	4525	4900	3892
Ag (relative)	1·00	0·95	1·55	3·41

MgI (Kersten and Ornstein 1941)

Initial level (ev.)	5·1	5·1	5·1	5·95	5·95	5·95	6·55	6·95
Wavelength (A.)	5184	5173	5167	3838	3832	3829	5528	4703
Ag (relative)	50	29·7	9·9	325	195	63·7	15	27·5

Excitation temperatures with barium and magnesium electrodes were calculated for equation (1) by substitution of the relative intensities determined and the appropriate Ag values, wavelengths and energy values.

With magnesium the mean result was of the order of 10,000° K., but in the case of barium and in determinations on the barium lines carried out with barium-magnesium electrodes the fluctuations were very much greater. The excitation temperature determined from the values obtained for the ionized barium lines was very much higher than that determined from values for the neutral barium lines. Table 2 summarizes these results, all of which were obtained with 110 amp. peak current. (For discussion see §6.)

Table 2

Excitation Temperatures (Ba lines)*

		Photographic results									E.M. results†	
Electrodes	Current pulse length (µsec.)	Ba II 4934/4900	Ba II 4934/4525	Ba I 6111/5778	Ba I 6063/5778	Ba I 6019/5778	Ba I 6111/5519	Ba I 6063/5519	Ba I 6019/5519	Ba I 5997/5519	Ba II 4934/4900	Ba II 4934/3892
Ba/Mg	1	11000	10000	5100	4500	3800						
				6000	5700	5060						
Ba/Mg	4	12000	13000	7400	5900	4700					12200	13400
				5400	5600	4400						
				7800		3900						
				9700	7500	4800						
Ba/Mg	9	13000	12200	6200	6200	5200						
				5900	5900	4800						
Ba	4	20000	22000	11000	8100	6400	12000	7000	5900	5500		
		24000	23000	11000	8000	7000	13000	8000	6800	6000		
		19000	19000	14000	9900	7600	10000	7200	6500	5700		

Excitation Temperatures (Mg lines)*

		Photographic results				E.M. results†
Electrodes	Current pulse length (µsec.)	Mg I 5184/5528	Mg I 5173/5528	Mg I 5184/4703	Mg I 5173/4703	Mg I 5184 (triplets) 3838 (triplets)
Mg	1	15000	11000	11000	9300	
		9300	8600	8400	8100	
Mg	4	12000	10000	9300	8700	7700
Mg	9	12000	10000	10000	9500	
		11000	9300	10000	9400	

* Wavelengths in A.

† E.M. = electron multiplier.

The calcium triplet λ4425, 4435, 4455 A., was measured to detect self-absorption since the records were exceptionally good. The ratios of the Ag values determined by Schuttevaer, de Bont and van den Broek (1943) give intensity ratios with ν^4 correction) 1 : 3 : 5·2 respectively, while from intensity determinations using the above technique, the ratios were 1 : 3·2 : 4 (again with ν^4 correction).

Further data showed that for the Cd triplet (4678, 4799, 5085 A.) the intensity ratios were 1 : 2·4 : 4·35 and 1 : 2·7 : 4·2 for two separate runs. For a Zn triplet (4680, 4722, 4811 A.) the ratios were 1 : 2·6 : 4·35. All these results include the ν^4 correction. The theoretical values are 1 : 3 : 5 and Ornstein, Hengstum and Brinkman (1938) obtained for the above Cd triplet an experimental value of 1 : 3·4 : 4·2 with an arc discharge. It is generally considered that ratios having values so far from the 1 : 3 : 5 value given by the Sum Rules indicate self-absorption for one or more of the lines. However, Mason (1938) found ratios of 1 : 3·32 : 5·37 for the Zn triplet and considered this result to be so near the theoretical value that self-absorption could be ignored. Mason and others point out that the Sum Rules are not necessarily valid for these wide multiplet separations, so that it seems reasonable to suppose that self-absorption was not important in our experiments. This is borne out to a first approximation by the general consistency of the temperatures given in table 2 for any particular set of conditions. Ornstein, Hengstum and Brinkman (1938) found ratios of 1 : 3·7 : 5·3 for the Cd $6\,^3S$–$5\,^3P$ triplet and, it is interesting to note, ratios of 0·98 : 2·92 : 5 for the multiplets $6\,^3D_{3,2,1}$–$5\,^3P_{2,1,0}$. A close triplet (Zn) with components at 3018, 3036, 3072 A. gave ratios of 0·99 : 3·13 : 5 (Schuttevaer and Smit 1943) which is extremely close to the Sum Rule value (see also Smith 1945); if the available resolution had been higher in the present experiments, this triplet intensity ratio would have been measured. Further experiments with ultra-violet lines will be undertaken. The results given in Table 2 are probably not very accurate since some self-absorption may be present, on the basis of the above discussion.

§ 5. EXPERIMENTAL RESULTS ON TRANSPORT OF METAL VAPOUR

A brief investigation of some electrode effects is described in order to emphasize the difficulty and approximations in interpreting the results of Table 2. The results described in this section have a wider bearing since they should also be of interest to spectroscopists engaged in analytical work.

Data on the transport of metal vapour in the discharge was obtained using two experimental methods: (*a*) observations on the spectrum of the electrode material in the discharge with an electron multiplier mounted in the focal plane of the spectrograph, and (*b*) photographing the discharge with a rotating mirror camera.

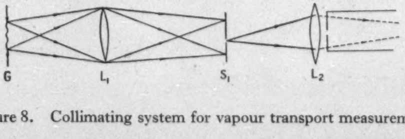
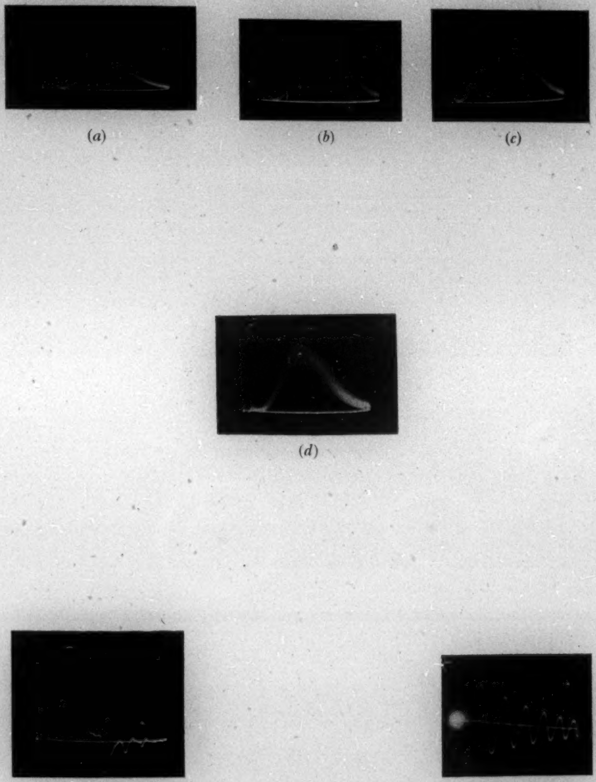


Figure 8. Collimating system for vapour transport measurements.

In the first of these methods, the intensity–time variation of a number of lines in the spectrum of cadmium, which was the electrode material used, was obtained for different regions along the length of the spark channel. Cadmium was used for these multiplier experiments on metal transport since other metals e.g. Zn, Ba and Mg showed such large spark-to-spark variations that the oscillograms were useless. In order that such measurements could be carried out, a collimating system as shown in Figure 8 was used. G is the spark gap, and an image of the

PROC. PHYS. SOC. VOL. 62, PT. 1—B (G. C. WILLIAMS, J. D. CRAGGS AND W. HOPWOOD)



(a) (b) (c)

(d)

(e) (f)

Figure 4.

To face page 56

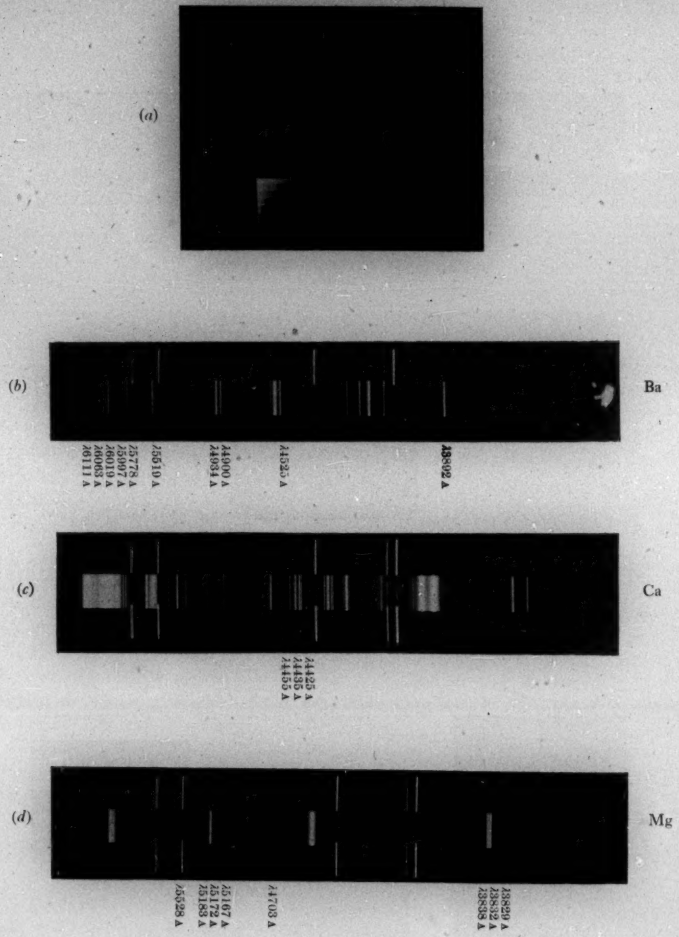

(a)
(b)
Ba
λ6111 Å
λ6063 Å
λ6019 Å
λ5997 Å
λ5778 Å
λ5519 Å
λ4934 Å
λ4900 Å
λ4525 Å
λ3892 Å
(c)
Ca
λ4455 Å
λ4435 Å
λ4425 Å
(d)
Mg
λ5528 Å
λ5183 Å
λ5172 Å
λ5167 Å
λ4703 Å
λ3838 Å
λ3832 Å
λ3829 Å

Figure 7.

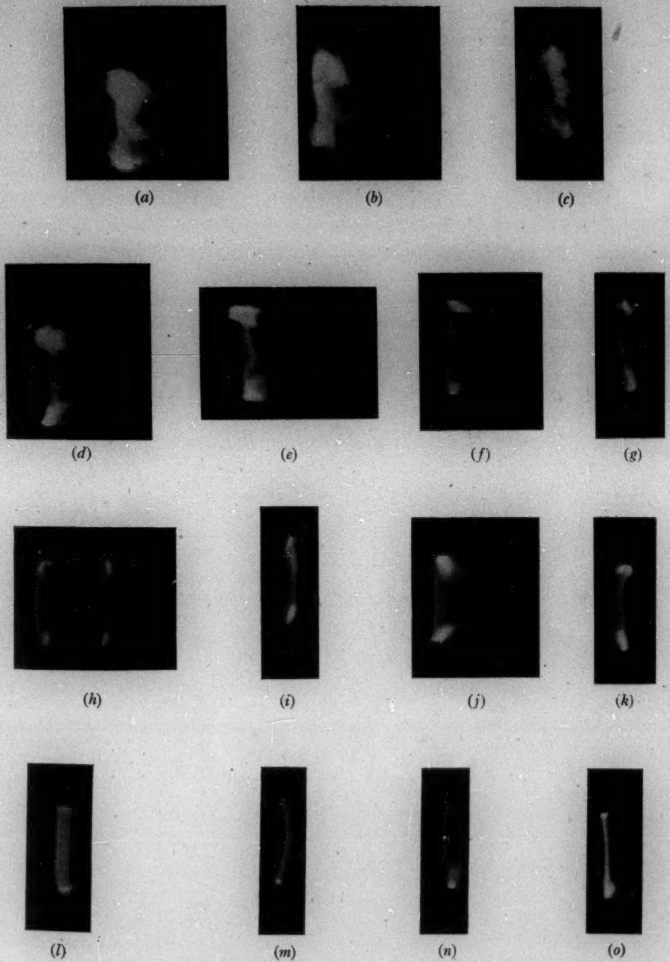


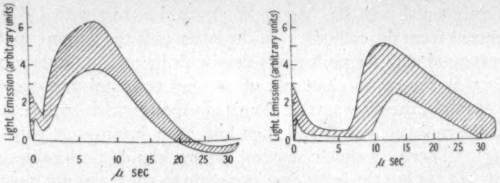
Figure 11. Rotating mirror and still records for sparks in air between Ba ((*a*), (*b*), (*c*) 9 μsec. sparks, (*d*), (*e*), (*f*), (*g*) 4 μsec. sparks), Mg ((*h*),(*i*) 4 μsec. sparks, (*j*), (*k*) 9 μsec. sparks), W ((*l*), (*m*) 9 μsec. sparks), and Cd ((*n*), (*o*) 4 μsec. sparks) electrodes.

spark was focused by means of the lens L_1 into the plane of the screen S_1. The screen carried a horizontal slit of variable width, and was mounted on a cathetometer, so that easy vertical movement and measurement of the slit position in the image was possible. Thus light from a small region of the image of the spark was allowed to pass through the collimating system into the spectrograph slit. An image ofthe slit S_1 was focused by means of the lens L_2, which was mounted directly before the spectrograph slit, on to the collimator lens. This was to ensure that all light which passed through the slit passed through the optical system of the spectrograph, and no false background was produced by reflection from the collimator walls.

In the experiments performed the length of the spark was about 6 mm. and was slightly magnified by the lens L_1 so that an image of length 8 mm. was produced. The width of the slit S was 0·1 mm., and observations of the light–time variation were made at intervals of 1 mm. along the length of the image.

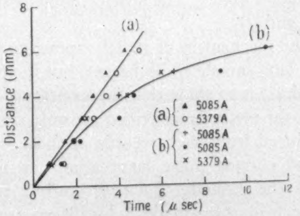
Measurements were carried out on the two Cd lines 5085 A. and 5379 A. for currents of about 120 amp. (duration 4 μsec.), and about 30 amp. (duration 10 μsec.). The observations were repeated a number of times over a period of weeks.

Figure 9 (*a*), (*b*) show typical oscillograms taken for a 10 μsec. spark at 1 mm. intervals along the spark image. The first part of the oscillograms, being slightly affected by pick-up (Figure 4 (*d*)), has been omitted. It can be seen that the onset of radiation takes place progressively later, and the radiation becomes weaker as the distance from the cathode increases. In some of the oscillograms it was observed that after having moved some distance along the channel the radiation became stronger again and occurred earlier in time, suggesting that some metal vapour was being given off at the anode. This was confirmed by the rotating



(*a*) 1 mm. from cathode. (*b*) 5 mm. from cathode.

Figure 9. Light/time oscillograms with cadmium electrode sparks.



(*a*) 4 μsec. current pulse.
(*b*) 10 μsec. current pulse.

Figure 10. Transport velocity of cadmium vapour in spark gaps.

mirror photographs; with Cd electrodes this anode radiation is much weaker than that originating at the cathode. The spark to spark variation in the intensity of the light originating from a small region of the discharge is much greater than that in the integrated light at any one wavelength.

It is possible from these results to determine the velocity of transport of the metal vapour through the gap (Figure 10 (*a*) and (*b*)), and it appears from the curve that the velocity is approximately $1{\cdot}1 \times 10^5$ cm/sec. for the 4 μsec. discharge, and that in the early parts of the 10 μsec. discharge (first 4 μsec.) the velocity is approximately the same, but falls steadily to about 4×10^4 cm/sec. after about 8 μsec.

Typical rotating mirror and still records for single sparks, using the same current waves as before, are shown for various arrangements in Figure 11. In all the photographs the anode is the upper electrode; the peak current in each case was 110 amp, the still records are Figure 11 (*c*), (*g*), (*i*), (*k*), (*m*) and (*o*).

There are many inexplicable features in the records of Figure 11. The extraordinary variation in the amount of material volatilized for each spark with Ba, which accounts for the great fluctuations in size of the electron multiplier light oscillograms, is to be compared with the steadiness of a discharge with Mg or Cd electrodes. The light emission from metal vapour emitted from either electrodes is approximately equal with Ba, Mg or W electrodes, but with Cd most of the vapour originates from the cathode. In the latter case the vapour speed is much higher (as measured with the multiplier) than with Ba or Mg photography, and it seems likely that several mechanisms of vapour emission are operative (§6). With Ba for example, there are several clouds of vapour which appear to move very rapidly into the mid-gap region and then lose their luminosity without further movement (*f*). There are slowly moving vapour clouds noticeable with Ba (*e*) and also with Mg (*j*), but the latter does not seem to show rapidly moving clouds. The afterglow in the channel (due possibly to nitrogen) is seen in (*l*), W electrodes, since the width of an infinitely thin 4 μsec. discharge should be only $\sim$1 mm. with the speed of mirror rotation used. The sharp channel can be seen on the left of the spark image in (*h*) and (*l*).

Further work on the mechanism of metal vapour emission in air and other gases is planned, using long and short discharges, but the present relevance of the records (Figures 10 and 11) is to show that the excitation temperatures as found for these discharges by the methods described are only time and space means.

Vapour speeds deduced from such records as those in Figure 11 for various electrode materials and current pulse lengths are as follows: Mg, 4 μsec., $2{\cdot}4 \times 10^4$ cm/sec. from the cathode, $2{\cdot}6 \times 10^4$ cm/sec. from the anode; Mg, 9 μsec., $2{\cdot}6 \times 10^3$ cm/sec. from the cathode, $1{\cdot}2 \times 10^4$ cm/sec. from the anode; Cd, 4 μsec., $2{\cdot}2 \times 10^4$ cm/sec. from the cathode; Ba, 9 μsec., $1{\cdot}3 \times 10^4$ cm/sec. from the cathode.

§6. DISCUSSION OF RESULTS

The excitation temperatures (Table 2) show that a statistical electron distribution, giving the same "temperature" for all levels, does not exist. Langstroth and McRae (1938 a, b) arrived at the same conclusion for fairly short sparks. The meaning of the temperatures of Table 2 needs clarification and reference should be made to the work of Ladenburg (1933) who points out that in cases where a thermal equilibrium does not exist a statistical equilibrium between excited atoms and electrons may be found. The electron velocity distribution should be Maxwellian and the excitation to and destruction of excited states must be due largely to electron impact.[*]

Ladenburg shows that the Klein-Rosseland (1921) relation

$$g_j\sqrt{E''}S_{jk}(E'')=g_k\sqrt{E'}S_{kj}(E'), \qquad \ldots\ldots(2)$$

may be applied to plasma conditions without the necessity for assuming a statistical equilibrium between electrons and excited atoms, since the S's and g's have values independent of an equilibrium; g_j and g_k are the statistical weights of the two energy levels, E_j and E_k. E'' and E' are the energies of the exciting electrons (before and after collisions) so that $E''>E_k-E_j$ and $E'=E''-(E_k-E_j)$. $S_{kj}(E')$ is the probability of destruction of an excited atom in a collision of the second kind with an electron. If the current density in the plasma is very high (as it is in spark channels), so that many such collisions occur, then the number of excited atoms destroyed per second is equal to the number produced, other processes being considered negligible; Ladenburg then finally finds that

$$N_k/N_j=(g_k/g_j)\exp\left[-(E_k-E_j)/\boldsymbol{k}T_{\mathrm{e}}\right], \qquad \ldots\ldots(3)$$

i.e. the ratio of populations of the two atomic levels corresponds to a statistical equilibrium at the electron temperature. The temperature of the normal atoms can be much lower than T_{e}. An excellent review of these and other matters was published by Llewellyn Jones (1945).

It appears from the results (Table 2) that the same T_{e} (about 10,000°) applies to the MgI levels studied, independently of current pulse length. For BaI, greater variations in T_{e} for the different levels are found and T_{e} (BaI) is greater for sparks between Ba electrodes than with Ba/Mg electrodes. T_{e} for the BaII levels is consistently higher than for BaI, reaching 20,000°. The relation between the nature of the electrodes and T_{e} is probably complicated and might be a function of electrode melting point, vapour pressure, and excitation level energies etc.

The persistence of luminosity of the metal vapour after the current had fallen to zero (Figures 5 and 11) is sufficient to prevent any information being obtained on the rate of loss of electronic energy in a currentless plasma at high pressure. The effect may be merely due to the finite lives of the excited states in question, but might be due to yet another complication, i.e. the excitation of metal atoms by collisions of the second kind with metastable atoms produced in the air during the earlier part of the discharge. The work of Manley and Duffendack (1935) and Duffendack and Thomson (1933) on line enhancement, e.g. of magnesium by neon, demonstrates the feasibility of this explanation. Bay and Steiner (1930) found, for example, that the metastable ($\mathrm{A}\,^3\Sigma$) nitrogen molecule could excite a mercury level ($2\,^3$S) at 7·7 volts. Further work with different gases will, it is hoped, give further information on this point, but the advantages of the new pulse techniques should enable the experiments to be simplified, and render them capable of adding new data to the earlier results of Duffendack and his collaborators.

---

[*] One of the referees of this paper has kindly brought to our notice the review article by Rompe and Steenbeck (1939) who discuss in detail "non-isothermal" plasmas. They point out (as have many other authors) that in many discharges, the electron, excitation and gas temperatures may be widely different although at higher pressures (10^{19} atoms/cm^3 or more) the electron temperature is only slightly less than the gas temperatures (say 20–100° less at 6000°) in steady maintained discharges (arcs). Rompe and Steenbeck do not treat transient conditions (cf. the work of Mannkopff and Paetz 1947). It seems reasonable to suggest, from the work of Ladenburg (1933) and of Mannkopff and Paetz that, in the sparks discussed in the present paper, thermal equilibrium will not be attained but that the excitation and electron temperatures will probably be almost equal (relative to a particular excited state) and that the several microseconds available during current flow should be enough for statistical equilibrium to be reached amongst the electrons. It seems that the excitation temperatures measured relative to the various energy levels are different (Table 2) and further work is planned to investigate this and other matters in more detail. Experiments to measure electron temperatures separately are also in progress.

The mean life (τ_n) of an excited state n is derived from a knowledge of Einstein A's since $\tau_n = 1/\Sigma A_{nm}$ where m is a lower state.

The A's must be absolute and not relative or as those given in Table 1, and such data seem to be lacking. Other methods are available for finding τ (Mitchell and Zemansky 1934, pp. 145 *et seq.*), but the τ values given are not for those states used in the present work. It is of interest to note that τ(Zn, 3076 A.) is ~10 μsec. and τ(Cd, 3261 A.) is ~2 μsec., so that true lives of the order of those shown for example in Figure 4 are not unreasonable.

It is assumed that in the time of current duration (1 μsec. or more) there have been sufficient collisions to enable the equilibrium between electron and excited atoms to be established. This seems reasonable in view of considerations involving the number of electronic collisions in 1 μsec. in spark channel conditions (Ladenburg 1933, Llewellyn Jones 1945, Margenau *et al.* 1946). This number of collisions is probably about 10^5, and the time required for the electrons in the channel to reach an equilibrium energy distribution should therefore be not more than about 1 μsec. The mean life of the excited states must be then much less than 1 μsec. for the Ladenburg equilibrium to be established, in which case the long afterglows could not be explained (see above) by the persistence of the excited atoms produced during the time of current flow.*

The mechanism of metal vapour emission at the electrodes may affect the values of T_e. Langstroth and McRae (1938 b) consider that diffusion or transport as charged particles could not account for the high vapour speeds (~10^5 cm/sec.) and suggest that an explosive process at the electrodes might be operative. Haynes (1946) in a short abstract has described in outline some work on the velocity of mercury vapour jets in hydrogen sparks of microsecond duration, using electron multipliers. Haynes concludes, without detailed analysis, that the high speed clouds or jets might be caused by the positive ions and electrons acquiring momentum in the anode and cathode drops respectively. This is an obvious general interpretation but it seems unlikely to hold for all the cases shown in Figure 11. The difference in behaviour for cadmium and magnesium electrodes is difficult to explain on such grounds, since (Mg, Figure 11 (*j*)) it is sometimes found that the anode and cathode clouds are very similar in appearance and speed of propagation. Llewellyn Jones (1946) has produced strong evidence and arguments in favour of a simple evaporation process, which is supported by independent work (Craggs, Haine and Meek 1947). In this case it is still difficult to account for the different behaviour of the various metals.

The movement of a cloud from the anode might be due to positive ions. Tyndall (1938, pp. 67 *et seq.*) shows that the drift velocity W of a positive ion is not proportional to X/p (field-strength/pressure) for values of $X/p > 10$ v/cm/mm. Hg. The mobilities for Mg^+, Cd^+, and Ba^+ in nitrogen are about 3, 2·3, and 2·2 cm/sec/v/cm. respectively at N.T.P. and the variation with temperatures up to 500° K. is negligible at constant density (Tyndall 1938, pp. 45, 58 *et seq.*). The initial gradient in the present spark gap is approximately 40 kv/cm. so that the ions mentioned could have drift velocities of approximately 8×10^5 cm/sec. which are adequate to account qualitatively for the speed of the anode vapour clouds. Further speculations, particularly on the cathode clouds, seem unjustified. The data on massive negative ions is extremely scanty.

---

* Mannkopff and Paetz (1947) have studied the relaxation time for the electron temperature in an ionized gas, and for atmospheric conditions find this time, i.e. the time for a Boltzmann equilibrium to be established, to be much less than 1 μsec.; they also state that at such high pressures the difference between electron and excitation temperature is negligible. More experimental data seem to be required and work to this end has been in progress at the University of Liverpool for some time.

§ 7. CONCLUSIONS

Excitation temperatures in 110-amp. sparks lasting 1–8 μsec. have been measured with various electrode metals and values varying between 3000° K. and 20,000° K. have been found. There is no clearly defined equilibrium between the electrons and excited atoms of all levels in the spark channels.

The results for these discharges, and those of the few earlier workers experimenting with similar discharges, must be interpreted with caution since the mechanisms by which the luminous vapour clouds are formed at the electrodes do not appear to be simple.

ACKNOWLEDGMENTS

The authors are grateful to Mr. F. R. Perry for his support of this work and to Sir Arthur P. M. Fleming, Director of Research and Education, and Mr. B. G. Churcher, Manager of Research Dept., Metropolitan-Vickers Electrical Co. Ltd., for permission to publish this paper.

REFERENCES

Bay, Z. and Steiner, W., 1930, *Z. phys. Chem.*, **9** B, 93.

Craggs, J. D., Haine, M. E., and Meek, J. M., 1947, *J. Instn. Elect. Engrs.*, Pt. III A, **93,** 963.

Craggs, J. D., and Meek, J. M., 1946, *Proc. Roy. Soc.* A, **186,** 241.

Dieke, G. H., Loh, H. V., and Crossewhite, H. M., 1946, *J. Opt. Soc. Amer.*, **36,** 185.

Duffendack, O. A., and Thomson, K., 1933, *Phys. Rev.*, **43,** 106.

Fitzgerald, M. A., and Sawyer, R. A., 1934, *Phys. Rev.*, **46,** 576.

Haynes, J. R., 1946, *Phys. Rev.*, **69,** 693.

Hopwood, W., and Craggs, J. D., 1947, *J. Opt. Soc. Amer.*, **37,** 560.

Jones, F. Llewellyn, 1945, *J. Soc. Chem. Ind.*, **64,** 317 ; 1946, *Nature, Lond.*, **157,** 298.

Kaiser, H., and Wallraff, A., 1939, *Ann. Phys., Lpz.*, **34,** 297.

Kersten, J. A. H., and Ornstein, L. S., 1941, *Physica*, **8,** 1124.

Klein, O., and Rosseland, S., 1921, *Z. Phys.*, **4,** 46.

Kruithof, A. M., 1943, *Physica*, **10,** 493.

Ladenburg, R., 1933, *Rev. Mod. Phys.*, **5,** 243.

Langstroth, G. O., and McRae, D. R., 1938 a, *Can. J. Res.* A, **16,** 17; 1938 b, *Ibid.*, **16,** 61.

Loeb, L. B., and Meek, J. M., 1940, *J. Appl. Phys.*, **11,** 438.

Manley, J. H., and Duffendack, O. S., 1935, *Phys. Rev.*, **47,** 56.

Mannkopff, R., and Paetz, H., 1947 (unpublished, see *Rev. Sci. Instrum.*, 1947, **18,** 142).

Margenau, H., McMillan, F. L., Dearnley, I. H., Pearsall, C. S., and Montgomery, C. G., 1946, *Phys. Rev.*, **70,** 349.

Mason, R. C., 1938, *Physica*, **5,** 777.

Mitchell, A. C. G., and Zemansky, M. W., 1934, *Resonance Radiation and Excited Atoms* (Cambridge : University Press).

Ornstein, L. S., and Brinkman, H., 1934, *Physica*, **1,** 797.

Ornstein, L. S., van Hengstum, J. P. A., and Brinkman, H., 1938, *Physica*, **5,** 145.

Rompe, R., and Steenbeck, M., 1939, *Erbebn. exakt. Naturw.*, **18,** 257.

Schuttevaer, J. W., de Bont, M. J., and van den Broek, T. H., 1943, *Physica*, **10,** 544.

Schuttevaer, J. W., and Smit, J. A., 1943, *Physica*, **10,** 502.

Smith, D. M., 1945, *Collected Papers on Metallurgical Analysis by the Spectrograph* (London : British Non-Ferrous Metals Association), p. 11.

Tyndall, A. M., 1938, *The Mobility of Positive Ions in Gases* (Cambridge : University Press).

Measurements on the Velocity of Sound in Mixtures of Hydrogen, Helium, Oxygen, Nitrogen and Carbon Monoxide at Low Temperatures

By A. van ITTERBEEK and W. van DONINCK
Physical Laboratory, University of Louvain, Belgium

Communicated by E. G. Richardson; MS. received 2nd June 1947,
in amended form 1*st August* 1948

§ 1. INTRODUCTION

In a previous paper (van Itterbeek and van Doninck 1946) we published a series of measurements on the velocity of sound in gas mixtures at low temperatures, the components being monatomic gases or hydrogen which possess no rotational energy at low temperatures.

In that paper we developed a theoretical method to compute the interaction term between the components, starting from measurements on the dependency of the velocity of sound as a function of pressure at fixed temperatures.

In the present paper we report on measurements of the same kind, but with components for which the rotational energy must be considered. So measurements are made on H_2–N_2, H_2–O_2, H_2–CO, He–N_2, He–O_2 between 90° K. and 65° K.

Just as for our previous measurements we find that the second virial coefficient B for the mixtures can be represented by means of the equation

$$B = B_1 x^2 + \beta x(1-x) + B_2(1-x)^2, \qquad \ldots\ldots(1)$$

B_1 and B_2 being the virial coefficients for the components. The mixtures H_2–N_2 are however an exception to this rule; in this case good agreement could only be obtained with the measurements by means of the following expression for B:

$$B = B_{N_2} x^{3/2} + \beta x^3(1-x) + B_{H_2}(1-x)^{3/2}. \qquad \ldots\ldots(2)$$

We suppose that the reason for this deviation is the great chemical affinity between the hydrogen and nitrogen molecules as it appears from the formation of ammonia. This fact led us to measure also the mixtures H_2–O_2 and H_2–CO. But as for the former mixtures, no deviations appear from equation (1).

The experimental arrangement has been described already, in part in the paper already referred to and in another paper in which this technique was first used (van Itterbeek and van Paemel 1938). It consists in essence of a quartz oscillator of frequency 523·78 kc/s. emitting ultrasonic radiation into a gas contained in a vessel in which the pressure can be varied. The movement of a reflector mounted on a micrometer screw enabled wavelengths in a gas to be measured by the reaction of the returning radiation falling on the emitter. The screw permitted measurements to one thousandth of a millimetre. The frequency of the quartz was measured at the boiling point of oxygen. A further measurement in liquid hydrogen indicated that variation of frequency of the quartz with temperature was not so important as possible changes in the temperature of the gas during a series, which constituted the greatest source of inaccuracy.

§ 2. THEORETICAL FORMULAE

The velocity of sound as a function of pressure in the mixture is given by the equation

$$W = W_0(1 + pS/RT), \qquad \ldots\ldots(3)$$

where W_0 is the velocity of sound in the gas mixture in the ideal state ($p = 0$), which can easily be computed from the concentrations and the specific heats of the components, R is the gas constant, T the absolute temperature, and

$$S = B + \frac{T}{\lambda}\frac{dB}{dT} + \frac{T^2}{2\lambda(\lambda+1)}\frac{d^2B}{dT^2}. \qquad \ldots\ldots(4)$$

λ is a numerical factor which can be computed from the specific heats and the concentration. S is computed by means of the method of least squares from the experimental curves represented analytically by equation (3).

For the case of the H_2–N_2 mixtures we found a good agreement with the experimental data by using the equation

$$S = S'_{N_2}x^{\frac{3}{2}} + \alpha x^3(1-x) + S'_{H_2}(1-x)^{\frac{3}{2}} \qquad \ldots\ldots(5)$$

and

$$S'_{N_2} = B_{N_2} + \frac{T}{\lambda_m}\frac{dB_{N_2}}{dT} + \frac{T^2}{2\lambda_m(\lambda_m+1)}\frac{d^2B_{N_2}}{dT^2}, \qquad \ldots\ldots(6)$$

and an analogous expression for S'_{H_2}.

Further

$$\alpha = \beta + \frac{T}{\lambda_m}\frac{d\beta}{dT} + \frac{T^2}{2\lambda_m(\lambda_m+1)}\frac{d^2\beta}{dT^2} \qquad \ldots\ldots(7)$$

and

$$\lambda_m = \lambda_1 x + \lambda_2(1-x)\,;\ \lambda_1 = (C_v^{(1)})_{p=0}/M_1.$$

For all the other mixtures we found

$$S = S_1'x^2 + \alpha x(1-x) + S_2'(1-x)^2. \qquad \ldots\ldots(8)$$

As for our previous measurements we found that the expressions

$$\alpha = [S - S_1'x^{\frac{3}{2}} - S_2'(1-x)^{\frac{3}{2}}] / x^3(1-x) \qquad \ldots\ldots(9)$$

and

$$\alpha = [S - S_1x^2 - S_2(1-x)^2] / x(1-x) \qquad \ldots\ldots(10)$$

are practically independent of concentration, which proves that the expression

$$\frac{T}{\lambda}\cdot\frac{d\beta}{dT} + \frac{T^2}{2\lambda(\lambda+1)}\cdot\frac{d^2\beta}{dT^2}$$

is small. Thus with good approximation we can write $\alpha = \beta$ and obtain some indication of the interaction between the components of the mixture.

(i) *Measurements on Hydrogen Mixtures*

(*a*) H_2–N_2. The velocity of sound as a function of pressure between 0·1 and 0·8 atmos. at liquid oxygen temperatures (90° K., 85° K., 80° K. and 75° K.) and for different concentrations has been measured. From these measurements we have computed the values of W_0 and S (see Figure 1, which gives typical data at 80° K., and Table 1). In this table we have compared the experimental values for W_0 with the values obtained theoretically for the mixture of the perfect gases.

In Table 2 are indicated the values for $\alpha=\beta$ computed by using equation (9). It appears that α is practically independent of concentration (cf. Figure 2).

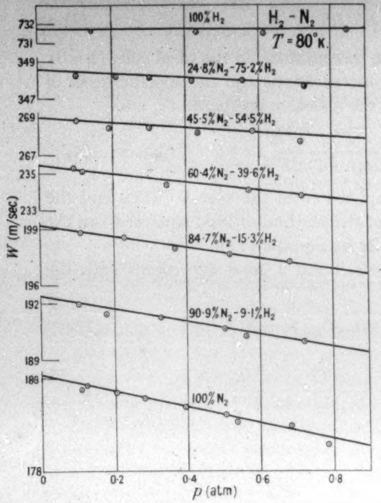


Figure 1.

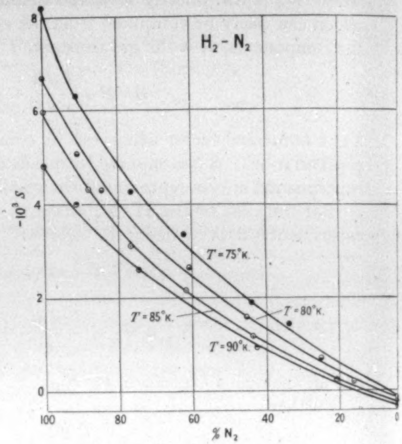


Figure 2.

(*b*) *H_2–O_2*. Similar experiments have been made on H_2–O_2 mixtures; the results of the calculations are indicated in Table 3 and Figure 3, and the values of α or β are computed by means of equation (10) and given in Table 4 and Figure 4.

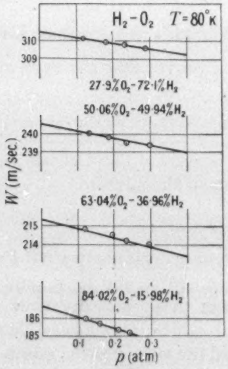


Figure 3.

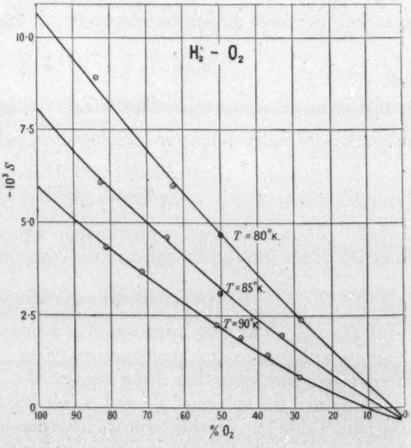


Figure 4.

Table 1 H_2–N_2 Mixtures

T (° K.)	Concentration %N_2	%H_2	W_0 (exp.) (m/sec.)	$10^3 \times S$	W_0 (theor.) (m/sec.)
90	100	0	193·8	−4·81	193·8
	91·3	8·7	205·5	−4·03	205·4
	74·6	25·4	224·5	−2·57	224·3
	42·8	57·2	291·7	−0·93	292·0
	21·2	78·8	392·1	−0·94	392·4
	0	100	773·0	0·29	
85	100	0	188·3	−5·96	188·1
	88·2	11·8	200·4	−4·31	200·2
	76·4	23·6	215·7	−2·94	215·8
	62·0	38·0	238·4	−2·16	238·2
	43·8	56·2	280·4	−1·17	280·4
	16·0	84·0	422·2	−0·23	422·1
	0	100	751·5	0·24	
80	100	0	183·1	−6·68	182·8
	90·9	9·1	192·2	−5·03	191·9
	84·7	15·3	199·0	−4·28	198·7
	60·4	39·6	235·2	−9·64	234·9
	45·5	54·5	268·9	−1·57	268·7
	24·8	75·2	348·4	−0·76	348·2
	0	100	731·4	0·17	
75	100	0	176·3	−8·21	176·6
	90·9	9·1	186·0	−6·32	185·7
	75·9	24·1	204·2	−4·24	204·3
	62·0	38·0	224·0	−3·37	224·1
	44·1	55·9	263·3	−1·91	263·0
	34·1	65·9	294·7	−1·43	294·7
	0	100	707·0	0·14	

Table 2

T (° K.)	x %N_2	$10^3 \times \alpha$	$10^3 \times \alpha$ average
85	90	5·5	6·0
	80	5·9	
	70	5·8	
	60	6·0	
	50	6·4	
	40	6·4	
80	90	8·1	8·0
	80	8·0	
	70	7·6	
	60	8·0	
	50	8·2	
	40	8·1	
75	90	9·0	9·6
	80	8·9	
	70	9·2	
	60	9·8	
	50	10·0	
	40	10·0	

Table 3 H_2–O_2 Mixtures

T (° K.)	Concentration %O_2	%H_2	W_0 (exp.) (m/sec.)	$10^3 \times S$	W_0 (theor.) (m/sec.)
90	100	0	181·0	− 5·90	
	81·2	18·8	200·2	− 4·31	200·1
	71·5	28·5	214·7	− 3.66	214·4
	50·8	49·2	253·2	− 2·26	253·1
	28·4	71·6	329·1	− 0·87	328·3
	0	100	773·0	0·29	
85	100	0	175·6	− 7·95	
	83·0	17·0	193·4	− 6·07	193·5
	64·7	35·3	219·0	− 4·59	219·1
	50·0	50·0	248·2	− 3·10	248·1
	33·3	66·7	299·2	− 1·97	299·0
	0	100	751·5	0·24	
80	100	0	171·0	−10·60	
	84·0	16·0	186·6	− 8·95	186·4
	63·1	36·9	215·2	− 5·98	215·3
	50·1	49·9	240·4	− 4·72	240·5
	27·9	72·1	310·4	− 2·40	310·5
	0	100	731·4	0·17	

Table 4

T (° K.)	x %O_2	$10^3 \times \alpha$	$10^3 \times \alpha$ average
90	80	−2·44	−2·9
	70	−2·83	
	60	−2·98	
	50	−3·05	
	40	−3·06	
	30	−3·06	
85	80	−4·78	−4·82
	70	−4·82	
	60	−4·83	
	50	−4·83	
	40	−4·83	
	30	−4·82	
80	80	−5·2	−6·5
	70	−6·0	
	60	−6·4	
	50	−7·2	
	40	−7·0	
	30	−7·2	

(*c*) *H_2–CO.* Figure 5 represents typical experimental data for the velocity; Figure 6 and Table 5 give the computed values for *S*.

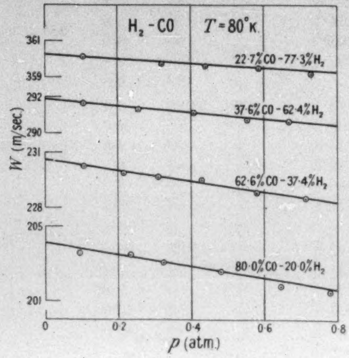


Figure 5.

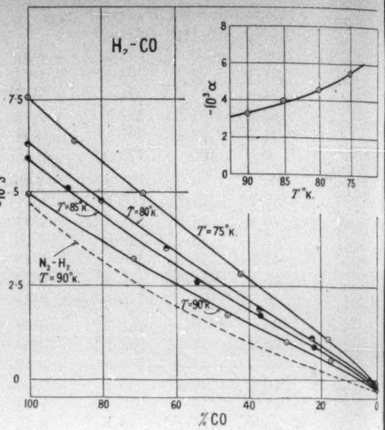


Figure 6.

Table 5. H_2–CO Measurements

T (° K.)	Concentration %CO	Concentration %H_2	W_0 (exp.) (m/sec.)	$10^3 \times S$	W_0 (theor.) (m/sec.)
90	100	0	193·7		
	71·2	28·8	229·6	−3·32	229·6
	46·3	53·7	282·1	−1·74	281·9
	29·7	70·3	342·7	−0·98	342·5
	17·3	82·7	423·3	−0·55	423·0
85	100	0	188·2	−5·92	188·0
	88·3	11·7	200·2	−5·10	200·4
	53·8	46·2	254·7	−2·59	255·0
	37·1	62·9	302·3	−1·70	302·3
	22·3	77·7	373·5	−0·84	373·7
80	100	0	183·0		
	80·0	20·0	204·1	−4·75	204·3
	62·6	37·4	230·6	−3·51	230·4
	37·6	62·4	291·9	−1·83	291·8
	22·7	77·3	360·3	−1·08	360·0
75	100	0	175·5		
	87·4	12·6	189·4	−6·37	189·3
	68·6	31·4	213·1	−5·00	213·3
	42·3	57·7	268·1	−2·82	268·0
	17·7	82·3	382·9	−1·07	383·0

For these measurements α and β could not be computed because the virial coefficients of CO are not known. We found however that *S* as a function of the concentration could be expressed by means of the equation

$$S = S_{\mathrm{CO}} x^2 + \alpha x(1-x) + S_{\mathrm{H}_2}(1-x^2).$$

S_{CO} was taken from our earlier measurements (van Itterbeek and van Doninck 1943). The values of α can however be considered as approximate values for β.

These are $10^3\alpha = -3{\cdot}29$, $-3{\cdot}96$, $-4{\cdot}26$, $-5{\cdot}43$ for $T = 90, 85, 80, 75$° K. respectively.

(ii) *Measurements on Helium Mixtures*

(*a*) *He–N$_2$*. Figure 7 gives W as a function of p and the concentration at 80° K. In Table 6 are indicated the values of S and in Table 7 the values of α or β (cf. Figure 8).

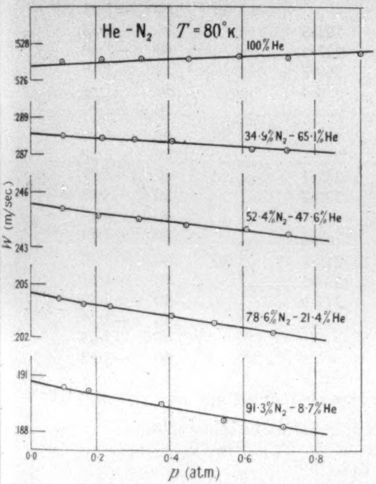


Figure 7.

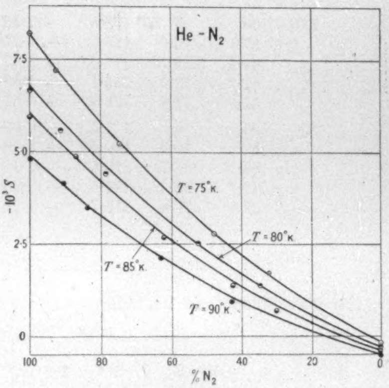


Figure 8.

Table 6

He–N$_2$ Mixtures

T (° K.)	Concentration %N$_2$	%He	W_0 (exp.) (m/sec.)	$10^3 \times S$	W_0 (theor.) (m/sec.)
90	90·4	9·6	203·4	−4·13	203·5
	85·9	14·1	210·4	−3·51	210·6
	62·7	37·3	240·9	−2·12	240·7
	42·8	57·2	281·9	−0·95	282·1
	0	100	559·5	0·50	559·2
85	87·3	12·7	200·6	−4·90	200·8
	61·7	38·3	235·7	−2·71	235·4
	42·4	57·6	275·4	−1·43	275·5
	29·6	70·4	315·4	−0·69	315·2
	0	100	542·9	0·38	542·7
80	91·3	8·7	190·8	−5·63	190·8
	78·6	21·4	204·5	−4·43	204·7
	52·4	47·6	245·2	−2·50	245·1
	34·9	65·1	288·2	−1·37	288·2
	0	100	526·9	0·28	527·1
75	92·6	7·4	183·5	−7·19	183·4
	74·7	25·3	202·7	−5·25	202·9
	47·6	52·4	246·8	−2·76	247·0
	32·2	67·8	286·7	−1·67	286·5
	0	100	509·9	0·19	510·2

Table 7

T (° K.)	x %N$_2$	$10^3 \times \alpha$	$10^3 \times \alpha$ average
90	80	−1·0	−1·2
	70	−1·0	
	60	−1·3	
	50	−1·4	
	40	−1·3	
	30	−1·3	
80	80	−2·6	−2·7
	70	−2·4	
	60	−2·5	
	50	−2·9	
	40	−2·9	
	30	−2·9	
75	80	−4·0	−4·0
	70	−3·7	
	60	−3·9	
	50	−4·0	
	40	−4·2	
	30	−4·2	

(*b*) *He–O$_2$*. Figure 9 represents typical experimental results of W as a function of pressure. Table 8 contains the values of S (Figure 10) computed from the measurements, and Table 9 values of α.

Table 8

He–O$_2$ Mixtures

T (° K.)	Concentration %O$_2$	%He	W_0 (exp.) (m/sec.)	$10^3 \times S$	W_0 (theor). (m/sec.)
90	100	0	181·0	− 5·90	
	89·1	10·9	191·8	− 4·94	191·6
	64·3	35·7	223·8	− 2·58	223·6
	41·7	58·3	269·8	− 1·12	269·7
	22·6	77·4	338·3	− 0·24	338·4
85	100	0	175·6	− 7·95	
	89·1	10·9	186·0	− 6·37	186·2
	64·3	35·7	217·4	− 3·82	217·1
	41·7	58·3	262·1	− 1·65	262·1
	22·6	77·4	328·9	− 0·61	326·7
80	100	0	171·0	−10·60	
	89·1	10·9	180·4	− 8·59	180·6
	64·3	35·7	210·6	− 5·36	210·8
	41·7	58·3	254·3	− 2·96	254·1
	22·6	77·4	318·7	− 1·21	319·0

Table 9

T (° K.)	x %O$_2$	$10^3 \times \alpha$	$10^3 \times \alpha$ average
90	80	−1·03	−1·5
	70	−1·33	
	60	−1·50	
	50	−1·56	
	40	−1·58	
	30	−2·00	
85	80	−1·99	−2·26
	70	−2·10	
	60	−2·29	
	50	−2·24	
	40	−2·42	
	30	−2·52	
80	80	−3·32	−3·34
	70	−3·42	
	60	−2·89	
	50	−3·24	
	40	−3·65	
	30	−3·52	

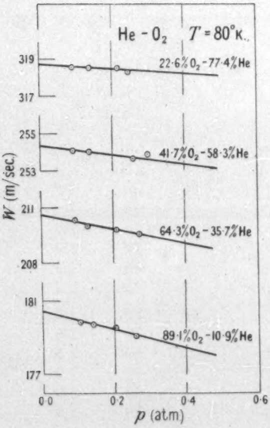


Figure 9.

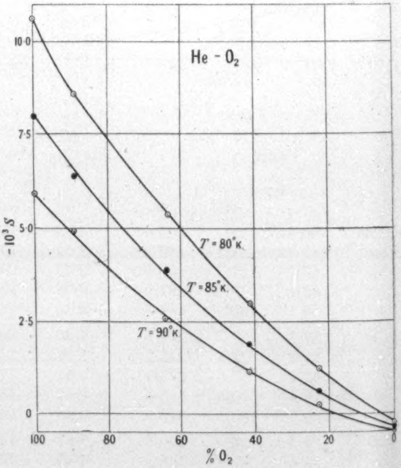


Figure 10.

§ 3. DISCUSSION

As we have said in the introduction, from our experimental results we can conclude that the second virial coefficient for the different mixtures can be represented by means of an equation of the type (1). The mixtures H$_2$–N$_2$ make an exception to this rule; for them other powers than the second must be chosen for the concentration.

In Table 10 we have summarized the different values of α together with the second virial coefficients for the components.

Table 10

Gas mixtures	T (° K.)	$10^3 \times a$	$10^3 \times B_1$	$10^3 \times B_2$
H_2–N_2	85	6·0	−0·3	− 9·3
	80	8·0	−0·4	−10·3
	75	9·6	−0·6	−11·6
H_2–O_2	90	−2·9	−0·2	− 9·3
	85	−4·8	−0·3	−10·0
	80	−6·5	−0·4	−11·0
H_2–CO	90	−3·3		
	85	−3·9		
	80	−4·6		
	75	−6·4		
He–N_2	90	−1·2	0·5	− 8·4
	80	−2·7	0·5	−10·3
	75	−4·0	0·5	−11·6
He–O_2	90	−1·5	0·5	− 9·3
	85	−2·26	0·5	−10·0
	80	−3·34	0·5	−11·0

From this table we see that there is a complete difference between the values obtained for the H_2–N_2 mixtures and the other ones.

REFERENCES

van Itterbeek, A., and van Doninck, W., 1943, *Physica*, **10**, 481; 1946, *Proc. Phys. Soc.*, **58**, 615.

van Itterbeek, A., and van Paemel, P., 1938, *Physica*, **5**, 593.

Ultrasonic Dispersion in Organic Liquids

By A. SCHALLAMACH
The British Rubber Producers' Research Association, Welwyn Garden City, Herts

MS. received 3*rd May* 1948

ABSTRACT. The sound velocities v have been measured, at frequencies f between about 700 and 3,000 kc/s., and at temperatures between about $+15°$ and $-80°$ c., of the following liquids : geraniol, di-dihydrocitronellyl ether, castor oil and iso-butyl alcohol. There is evidence of a positive dispersion ($\partial v/\partial f > 0$) in the first three liquids at low temperatures. A negative dispersion has been observed in geraniol and di-dihydrocitronellyl ether at higher temperatures, and in iso-butyl alcohol over the whole temperature range.

The positive dispersions are discussed on the basis of the classical hydrodynamic theory, and the implications of a complex viscosity are examined.

---

§ 1. INTRODUCTION

In classical hydrodynamics, the amplitude of a plane sound wave in a viscous liquid is given by (Lamb 1945)

$$u = u_0 \exp(j\omega t + mx), \qquad \ldots\ldots(1)$$

with

$$m = -j\omega/\sqrt{(\kappa/\rho)}\,(1 + 4j\omega\eta/3\kappa)^{-\frac{1}{2}}, \qquad \ldots\ldots(2)$$

where x is the direction of propagation, and ω the pulsatance of the wave in a liquid of density ρ, viscosity η, and adiabatic bulk modulus κ. The sound velocity is found from the relation

$$1/v = -I(m)/\omega. \qquad \ldots\ldots(3)$$

If $\omega\eta/\kappa \ll 1$, one has, on expanding equation (2)

$$v = v_0 \Big/ \left[1 - \frac{2}{3}\left(\frac{\omega\eta}{\kappa}\right)^2 - \frac{70}{729}\left(\frac{\omega\eta}{\kappa}\right)^4 - \ldots\right] \simeq v_0\left[1 + \frac{2}{3}\left(\frac{\omega\eta}{\kappa}\right)^2\right] \qquad \ldots\ldots(4)$$

where

$$v_0 = \sqrt{(\kappa/\rho)}. \qquad \ldots\ldots(5)$$

Equation (4) shows that the classical theory predicts a dispersion of the sound velocity if the product of frequency and viscosity becomes comparable with the bulk modulus. Ordinarily, this effect is negligible and the standard text-books give therefore equation (5) for the sound velocity.

Apart from the viscosity effect, no dispersion should be found as long as the mechanical properties of the liquid do not depend on the frequency. The bulk of the experimental evidence indicates that there is no ultrasonic dispersion in liquids up to about 100 Mc/s. (Parthasarathy 1936, Bär 1938, Krishnan 1939, 1941, Bhagavantam and Rao 1946). Exceptions are castor oil for which Zachoval (1939) reported a rise of the sound velocity with increasing frequency, and carbon disulphide where, according to Te-Tchao (1946), the velocity drops with increasing frequency.

The sound velocity measurements at low temperatures reported in this paper have been carried out in certain long-chain liquids which do not crystallize at these temperatures but eventually set to a glass. It was thought that the continuous transition to a glass might affect the ultrasonic properties of these liquids even at only moderately high frequencies. At the same time, regions of temperature and frequency where dielectric relaxation occurs would be covered in these experiments, so that it would be possible to compare these two effects and to see if there were any connection between them.

§ 2. EXPERIMENTAL

Two suitable liquids used in this investigation are geraniol

$$(CH_3)_2 . C : CH . (CH_2)_2 . C(CH_3) : CH . CH_2 . OH$$

and di-dihydrocitronellyl ether

$$[(CH_3)_2 . CH . (CH_2)_3 . CH(CH_3) . (CH_2)_2]_2 . O.$$

The relevant data for their purity were given in an earlier publication (Schallamach 1946). Since a preliminary note on this work appeared (Schallamach 1948) we experimented, in addition to these long chain liquids, with castor oil of medicinal quality "cold drawn". For comparison, the sound velocities in iso-butyl alcohol were determined. The sample used was a commercial product.

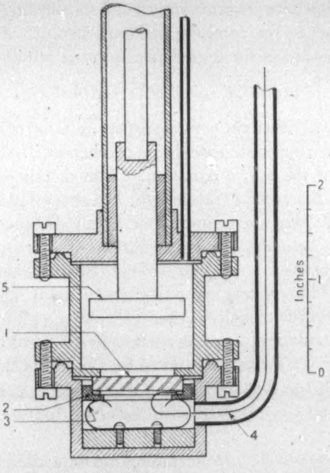


Figure 1. The interferometer. 1, crystal; 2, ebonite holder; 3, spring; 4, high tension lead; 5, reflector.

The measurements were carried out in a sonic interferometer specially adapted for low temperatures. The essential part of this instrument is shown in Figure 1. The X-cut circular quartz with silvered electrodes was pressed against the rim of an orifice in the bottom of the liquid container, without an intervening membrane. In this way, good sound transmission was assured at all temperatures. In the course of two days, ample time for an experiment, a few drops oozed out of the vessel but this did not affect the measurements. As no packing material suggested itself for the other joints, metal-to-metal contacts were relied upon for tightness: the bold rings on one set of flanges pressed against a flat part of the other flanges served this purpose well. The reflector was adjusted by a micrometer arrangement about 10 inches above the interferometer vessel which was immersed in a well stirred acetone-Drikold bath and maintained at an internal nitrogen pressure of about 10 inches of water. The temperature was measured by a thermocouple in contact with the interferometer and was accurate to ±0·1° c.

The quartz crystal was wired in parallel with a resonant circuit which was loosely coupled to an oscillator, and the voltage across the crystal was measured by a valve voltmeter. It is easily seen that this voltage indicates the equivalent impedance of the system crystal–liquid column. Hubbard (1931) has shown that this impedance consists of two series resonant circuits, connected again in series. One of the resonant circuits represents the crystal, and the other, representing the liquid column, is found to be formally identical with the input impedance of a transmission line. As long as the damping of the sound waves in the liquid is not too great the impedance has a minimum whenever the position of the reflector coincides with a node of the particle velocity. As the reflector is moved through the liquid sharp decreases in voltage across the crystal are observed at positions of the reflector half a wavelength apart.

This simple relation between wavelength and minimum of the impedance is no longer strictly valid when the damping becomes strong. It is then necessary to evaluate the exact condition for impedance minima which is

$$\tanh(2\alpha l) = -(2\pi/\lambda\alpha)\tan(4\pi l/\lambda),$$

where l is the distance between the crystal surface and the reflector, λ the wavelength of the sound and α the amplitude absorption coefficient. Abadie (1946) has given graphical solutions of the above equation. The distances between impedance minima are now less than half a wavelength, but this effect is serious only for the first half-wave and quickly loses importance as the distance between quartz and reflector increases. Actually, the construction of our interferometer did not allow measurements of the first few half-waves to be made and the error is small for this reason alone. The greatest error likely to arise from this source has been estimated for di-dihydrocitronellyl ether at −80° c. The viscosity is known (Schallamach 1946), and the absorption coefficient α was calculated by means of the classical equation $\alpha = R(m)$ where m is given by equation (2). It was found that at 2,102 kc/s. the error is less than 0·01%, and that at 689 kc/s. less than 0·1%. This value is, at −80° c., small compared with the error due to other sources and was therefore neglected.

The accuracy of the wavelength determination depended mainly on the distance over which a standing wave could be established with certainty, and it became progressively worse as the temperature fell and the attenuation increased. Near room temperature where the damping is unimportant the wavelengths were accurate to ±0·05%, and this accuracy was maintained in the long chain liquids down to about −60° c. where the error began to rise until it reached ±0·35% at about −80° c.

The crystals were excited in their fundamental frequency and its third harmonic. The highest frequency used was approximately 3,000 kc/s. but most measurements were carried out at about 700 and 2,000 kc/s. Using a heterodyne wavemeter the frequency could be measured to an accuracy within ±0·05%. Thus, the sound velocity determined as a product of frequency and wavelength was accurate to between ±0·1% to ±0·4%.

As a check on the apparatus, the sound velocity in water was determined at two frequencies. Water was chosen for this purpose because it represents the best established case of a liquid with no dispersion. The values obtained (in m/sec.) were:

Temperature ° c.		23	16·6
Frequency (kc/s.)	689	1489	1468
	2102	1491	1470
Willard (1947)		1493	1476

This table shows that there is agreement between the measurements at different frequencies, and reasonable agreement with the values obtained from an empirical equation given by Willard.

§ 3. RESULTS

Figures 2 and 3 give the results of the last of a number of experiments with geraniol and di-dihydrocitronellyl ether; improvement in technique during these experiments did not materially affect the results. As will be seen, both graphs show the same type of curves: at any one frequency the sound velocity increases with decreasing temperature. This increase is almost linear at higher temperatures, and in the same region the sound velocity has a small negative frequency coefficient. For example, comparing measurements in geraniol at 689 and 2,102 kc/s., the relative change at 10° c. is 0·2%, which is near the experimental error, but at -50° c. it increases to 0·65%. From the number of points which give a consistent result, this effect appears to be real and will be referred to as negative dispersion. At lower temperatures, the temperature coefficient becomes considerably larger and the curves cross each other so that the frequency coefficient changes its sign. This positive dispersion increases as the temperature falls and

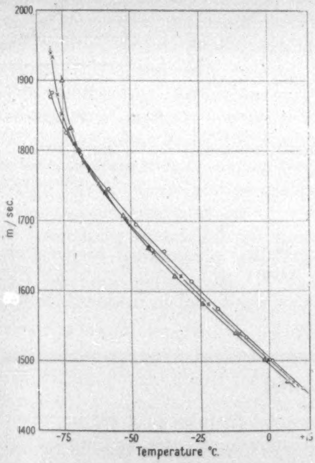


Figure 2. Sound velocity in geraniol, *vs* temperature. ○ 689 kc/s.; × 2,102 kc/s.; Δ 2,943 kc/s.

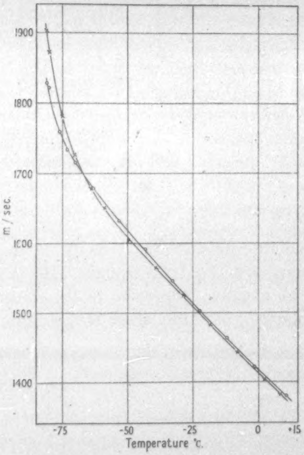


Figure 3. Sound velocity in di-dihydrocitronellyl ether, *vs* temperature. ○ 689 kc/s. ; × 2,102 kc/s.

at −80° c. the relative difference between 689 and 2,102 kc/s. is about 3% in both liquids. We should mention here that extending the temperature range would give little advantage as the damping would make accurate measurements impossible at still lower temperatures.

There is only evidence of a positive dispersion in castor oil (Figure 4), but at rather lower temperatures than reported by Zachoval who found an increase of 1% at 18° c. when increasing the frequency from about 2,000 to 6,000 kc/s., and the absolute value of the velocity at this temperature is somewhat higher than Zachoval's, viz. 1,529 against 1,496 m/sec. at 2,000 kc/s.

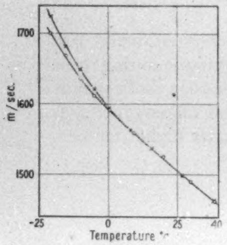


Figure 4. Sound velocity in castor oil, *vs* temperature. ○ 689 kc/s. ; × 2,102 kc/s.

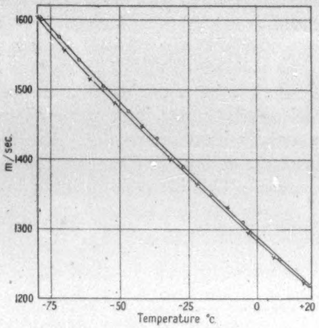


Figure 5. Sound velocity in iso-butyl alcohol, *vs* temperature. ○ 689 kc/s.; × 2,102 kc/s.

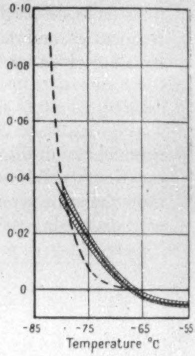


Figure 6. Relative difference of the sound velocity in di-dihydrocitronellyl ether at 689 and 2,102 kc/s.
—— experimental
- - - theoretical

A negative disperison of the sound velocity was found in iso-butyl alcohol (Figure 5) over the whole temperature range.

§ 4. DISCUSSION

A negative temperature coefficient, at constant frequency, of the ultrasonic velocity of the type observed at higher temperature in all cases, and in iso-butyl alcohol at all temperatures within the experimental region, conforms with the results generally obtained with liquids other than water (Freyer, Hubbard and Andrews 1929, Prosorov and Nozdrev 1939, Suryaprakasan 1940, Willard 1947).

The negative dispersion found at these temperatures, as shown in Figures 2, 3 and 5, is small but it is nevertheless well substantiated, as pointed out above. This effect was quite unexpected and appears very difficult to account for. The only reference to the possibility of a negative dispersion is a theoretical paper by Ginsburg (1942) who bases his arguments on certain modifications of the Navier-Stokes equations, but the order of magnitude of the effect as predicted by this theory is much smaller than the values observed here.

There is no singularity in the curves at temperatures and frequencies where dielectric relaxation occurs. The following table gives the temperatures in degrees centigrade at which the loss factor has a maximum:

Frequency (kc/s.)	2943	2102	689
Geraniol *	−39°	−43°	−51°
Di-dihydrocitronellyl ether *		−66°	−72°
Iso-butyl alcohol †		−65°	−81°

* Schallamach 1946. † Kobeko, Kuvshinsky and Shishkin 1938.

It will be seen that there is no obvious connection between the two processes.

The sharp rise of the sound velocity at lower temperatures appears to begin near the temperature where the curves cross and the positive dispersion starts. A similar increase has been observed in liquid oxygen below −200° c. by Liepman (1936) who, however, worked only at one frequency. In the case of our liquids it may well be that this effect is connected with the transition to a glass referred to in § 1. One would naturally expect a much higher sound velocity in the glassy state but it is surprising that the transition should make itself felt at comparatively high temperatures because a liquid like di-dihydrocitronellyl ether is brittle only in the neighbourhood of −140° c. Before discussing the possibility of frequency dependent mechanical constants, which is further suggested by the observed positive dispersion, the implications of the classical theory must be examined. The condition for a noticeable positive dispersion as predicted by hydrodynamics is fulfilled here because the quotient $(\omega\eta/\kappa)^2$ can no longer be neglected in comparison with unity. An analysis of the data is possible for the ether as the viscosities are known at the temperatures in question (Schallamach 1946). A certain complication arises from the negative dispersion which makes the difference between the velocities at the two frequencies ambiguous for this purpose. The relative differences are plotted against temperature in Figure 6, the shaded area around this curve indicates the experimental uncertainty, and the broken line gives the theoretically expected differences calculated from equation (4), neglecting the very small correction at the lower frequency. It will be seen that there is rough agreement between the two curves but it appears that the experimental curve begins to rise at a rather higher temperature than the theoretical curve, and that it falls below it at lower temperatures. This latter effect is more pronounced in castor oil for which the correction has been calculated with a viscosity extrapolated from data in the International Critical Tables. At −20·8° c., the observed relative velocity change is 1·5% and the theoretical difference 4·6%.

The experimental results acquire a different aspect if the liquid is taken to be a Maxwellian body characterized by a mechanical relaxation time τ and a complex viscosity

$$\eta^* = \eta_0(1 - j\omega\tau)/(1 + \omega^2\tau^2). \qquad \ldots\ldots(6)$$

Such a liquid has rigidity against alternating stresses the period of which is comparable with the relaxation time. Introduction of a complex viscosity into equation (2) leads to expression (7) for the sound velocity which is correct to the same degree of approximation as equation (4):

$$v = v_0\left\{1 + \frac{2}{3}\left(\frac{\eta_0\omega}{\kappa}\right)^2\left[\frac{\kappa}{\mu}\,\frac{1}{1 + \omega^2\tau^2} + \frac{1 - \omega^2\tau^2}{(1 + \omega^2\tau^2)^2}\right]\right\}, \qquad \ldots\ldots(7)$$

where $\mu = \eta_0/\tau$ is the rigidity at infinitely high frequency. The second term in square brackets corresponds to the viscosity dispersion, which is now smaller than in an ideal liquid, but the first term gives the increase of the velocity due to the rigidity. It is difficult to estimate the value of the limiting rigidity μ but the order of magnitude of the ratio κ/μ is approximately 3 for ordinary solids, and we feel justified in assuming that μ is, in any case, considerably smaller than κ. Equation (7) shows that the effective rigidity produces a greater dispersion than the viscosity. If, therefore, the conceptions underlying the formulation of the complex viscosity are accepted, it follows that any observed dispersion must be primarily due to the rigidity of the liquid.

The experimental evidence is insufficient to determine whether the dispersion reported here may be asscribed to complex viscosity. Measurements of the absorption coefficient would most probably give an answer to this question.

It may be argued that it is not necessary to restrict the discussion to a frequency-dependent viscosity but that similar considerations could be applied to the compressibility. It must, however, be borne in mind that this would involve the introduction of a compressional or " second " viscosity the existence of which is not recognized in hydrodynamics and which has by no means been proved yet experimentally.

ACKNOWLEDGMENT

This work forms part of a programme of fundamental research undertaken by the Board of the British Rubber Producers' Research Association; the sonic interferometer described above was made by Mr. H. W. Giles in the workshop of the Association.

REFERENCES

ABADIE, P., 1946, *Trans. Faraday Soc.*, **42**A, 143.

BÄR, R., 1938, *Proc. Ind. Acad. Sci.* A, **8**, 289.

BHAGAVANTAM, S., and RAO, B. R., 1946, *Nature, Lond.*, **158**, 484.

FREYER, E. B., HUBBARD, J. C., and ANDREWS, D. H., 1929, *J. Amer. Chem. Soc.*, **51**, 759.

GINSBURG, V. L., 1942, *C.R. Acad. Sci. U.S.S.R.*, **36**, 8.

HUBBARD, J. C., 1931, *Phys. Rev.*, **38**, 1011.

KOBEKO, P., KUVSHINSKY, E., and SHISHKIN, N., 1938, *Tech. Phys. U.S.S.R.*, **5**, 413.

KRISHNAN, K. G., 1939, *Proc. Ind. Acad. Sci.* A, **9**, 382; 1941, *Ibid.*, **13**, 281.

LAMB, H., 1945, *Hydrodynamics* (New York : Dover Publications).

LIEPMAN, H. W., 1936, *Helv. Phys. Acta*, **9**, 507.

PARTHASARATHY, S., 1936, *Proc. Ind. Acad. Sci.* A, **4**, 17.

PROSOROV, P., and NOZDREV, V., 1939, *J. Exp. Theor. Phys. U.S.S.R.*, **9**, 625.

SCHALLAMACH, A., 1946, *Trans. Faraday Soc.*, **42**, 495; 1948, *Nature, Lond.*, **161**, 476.

SURYAPRAKASAN, V., 1940, *Proc. Ind. Acad. Sci.* A, **12**, 341.

TE-TCHAO, O., 1946, *C.R. Acad. Sci., Paris*, **222**, 1165.

WILLARD, G. W., 1947, *J. Acoust. Soc. Amer.*, **19**, 235.

ZACHOVAL, L., 1939, *C.R. Acad. Sci., Paris*, **208**, 265.

REVIEWS OF BOOKS

High Vacua, by Swami Jnanananda. Pp. xiii+310. (New York: D. Van Nostrand Co. Inc., 1947.) 30*s*.

It is a welcome feature of this book that prominence is given to the Kinetic Gas Theory, without a sure understanding of which it is impossible to understand many aspects of high vacuum technique. The author has, in fact, devoted one quarter of the book to this most important subject, which is treated under the following sectional headings : Introduction, Free Path Phenomena, Transport Phenomena, The Equation of State, Free Molecular Behaviour of Gases. The next quarter of the book is given up to a description of vacuum pumps, amongst them being " Pistonless High-Vacuum Pumps with Oils or Paraffins or High-Boiling Phlegmatic Liquids as Pumping Media ", i.e. Oil Diffusion Pumps. Then follows a long chapter giving descriptions of many types of gauges for pressure measurement, and the remaining three short chapters are devoted to the Technique of High Vacuum, Preparatory Operations for Vacuum Work, and finally, Production of High Vacua by Physical-chemical Methods. The book is completed by the insertion of three short tables and an index.

In appraising a book it is natural to ask why the book was written and what purpose it was intended to serve. The answer to these questions is given in the last paragraph of the Foreword by George A. Lindsay : " The work supplies a real need for collected and organized information on the technique of the production of high vacua and should be a valuable reference for the worker in that field ". It is true to say, particularly with regard to the second or third chapters on the production and measurement of low pressures respectively, which together form over half the book, that the author has carried out an immense amount of solid work in presenting an exhaustive catalogue of pumps and gauges, and from this point of view the aim of supplying a " reference work " has been achieved. But it is also pertinent to ask whether at this stage the worker in the high vacuum field requires a historical reference work or rather an up-to-date manual of vacuum practice, for it is quite certain that if a worker is looking for the latter he will not find it in this book. It is perhaps the reviewer's chief criticism that, whereas during the war there has been a phenomenal advance in the science of high vacuum technique, there is no reference in the book to any work published later than 1940, and for this very reason the reviewer experienced an acute feeling of disappointment, which tended to overshadow the appreciation of the author's work in compiling a history of the subject. This feeling of disappointment was unfortunately in no way lessened when a perusal was made of the manner in which the author dealt with certain aspects which are treated in most text-books on the subject. The reviewer has often found that in earlier works the explanation of certain features has been far from lucid, and, if one may be permitted to say so, this tradition has been well maintained in the present book.

For example, the essential difference between Gaede's " rate of exhaustion ", and Langmuir's " speed of a pump " does not appear to be brought out at all in the explanation given on pp. 75, 76 where S has to do duty for both in equations (2–2 *a*) and (2–3 *b*). It is not very helpful to define the speed, as on p. 74, as the volume or mass of gas abstracted per unit time at a given pressure.

Again, the discussion of conductivity and the flow of gases through pipes is extremely confusing. On pp. 77, 78 F is used to denote " conductivity " or rate of flow, and Q the volume of gas flowing out, measured at a pressure of 1 dyne/cm^2. This should be compared with the definition on p. 103 for the output Q, viz. the " product of the mean pressure and the volume removed per second, under a pressure difference of 1 dyne/cm^2 ". In this case Q is the conductivity which was earlier denoted by F. On p. 228, Q is defined as " the quantity of gas flowing per sec." This is, of course, the best explanation, assuming that it is understood that " quantity " refers to " pressure × volume ", as indicated on p. 59. On p. 79, however, Q is defined as " the quantity of gas taken out by the pump ". The confusion of thought is also apparent on p. 64 when F is incorrectly described as the rate of flow in cm^3/sec. per unit pressure difference, and again on p. 78 in the caption to figure 2-1. In transforming equation (1–129 *a*) to (1–129 *b*) the author has taken into account end effects only for the case of molecular flow, and has ignored them for the viscous flow. A comparison of equations (1–126 *b*), which is correct, and (2–6), shows the presence of an algebraical error, and this affects equations (1–129 *b*) and (1–129 *c*) equally.

The author follows the usual procedure when he says on p. 115 that the speed of Langmuir's pump was fifty times as great as that of Gaede's pump inferring that 80 cm^3/sec. was the maximum speed to be obtained from a diffusion pump, and omitting any reference to Gaede's later pumps described by him in 1923.

Every reader may not agree with some of the statements made in the text. For example, it does not appear *necessary* in order to measure the speeds of pumps over an appreciable pressure range to use any other method than that described on p. 75. In some cases, however, the method described on p. 76 may be used as an *alternative*. Again, on p. 111, it is stated that in a diffusion pump, the gas molecules diffuse into the stream of mercury vapour, " as the partial pressure of the gas is higher than that of mercury vapour ", and on p. 112, that " the optimum diffusion of gas molecules takes place only when the width of the aperture is comparable with their mean free path ".

It is not the universally accepted way of sealing demountable joints to coat them both with a thin film of grease and then join them gently but firmly, thus filling the space between the surfaces with grease (p. 250) ; further, graphite is not a constituent of " Apiezon Q ".

There appear to be very few printer's errors, although one or two have been observed, e.g. on p. 66 equation (1–129 *a*) should have the term $D^3/2{\cdot}394L$; on p. 76 equation (2–4) should read $\log_e [(p_1-p_0)/(p_2-p_0)]$; on p. 203, the equivalent of 0·4 mil. should read 0·001 cm., and a similar correction is required on the top line of p. 204 ; on p. 217, equation (3–22) should read $dr+dcf(p)$.

R. WITTY.

Technical Optics, Volume 1, by L. C. MARTIN. Pp. vii + 393. (London: Sir Isaac Pitman & Sons, Ltd., 1948.) 40*s*. net.

In the present dearth of books on optics the appearance of the first volume of a revised, enlarged (and renamed) edition of Professor Martin's *Introduction to Applied Optics* is more than welcome. The scope of this work, as that of its forbear, is wide—almost too wide for anything like completeness to be attempted, although this disadvantage is offset by the very generous list of references at the end of each chapter.

The first four chapters treat the paraxial theory of lens systems, the wave theory of light (including an introduction to the diffraction theory of aberrations), and the aberrations of symmetrical systems. The later chapters provide self-contained introductions to the optics of vision, physical optics and doubly refracting crystals, the elements of the theory of radiation and dispersion, and finally chapters dealing with the working and testing of optical surfaces and with the theory of spectacles.

In a work embracing such a wide field it is gratifying that a careful reading only revealed one error of consequence. In dealing with the theory of Petzval curvature (p. 137) the usual expression $\left[\frac{1}{r}\left(\frac{1}{N}-\frac{1}{N'}\right)\right]$ is obtained, but it is stated that this result is not valid in the case when the object is at the centre of curvature of the refracting surface. The proof given earlier (p. 23) in support of this contention refers to the curvature of the surface containing the tangential focus and not to the Petzval surface. In fact the usual form of the expression for the Petzval curvature remains valid no matter what the position of the object. A proof valid in this special case is given in Whittaker's *Theory of Optical Instruments* (Cambridge Mathematical Tracts), although the proof given there breaks down when the object and the refracting surface coincide. It is a surprising fact that no *universally* valid proof of this well-known theorem ever appears to have been published.

Apart from this, one can have nothing but praise for this book. One can scarcely say the same for the price. This, following the general upward trend, will no doubt make many students think twice about buying a book which they ought to buy without any hesitation.

H. H. HOPKINS.

CONTENTS FOR SECTION A

PAGE
Editorial 1
Dr. J. HAMILTON. Collision Problems and the Theory of Radiation Damping . 2
Dr. J. HAMILTON. Damping Theory and the Propagation of Radiation . . 12
Dr. M. KROOK. Continuous γ-Emission in Neutron–Proton Collisions . . 19
Dr. E. J. BOWEN, Mr. E. MIKIEWICZ and Mr. F. W. SMITH. Resonance Transfer of Electronic Energy in Organic Crystals 26
Mr. A. G. FENTON and Mr. E. W. FULLER. Further Experiments with an Adjustable Geiger-Müller Counter 32
Dr. R. EISENSCHITZ. The Steady Non-Uniform State for a Liquid . . . 41
Dr. A. H. COTTRELL and Mr. B. A. BILBY. Dislocation Theory of Yielding and Strain Ageing of Iron 49
Mr. R. HUBY. Investigations on the Binding Energy of Heavy Nuclei . . . 62
Reviews of Books 71
Contents for Section B 75
Abstracts for Section B 75

ABSTRACTS FOR SECTION A

Collision Problems and the Theory of Radiation Damping, by J. HAMILTON.

ABSTRACT. The relation between radiation damping theory and the problem of the radiationless collision of an electron with a field of force is discussed. It is shown that the divergence difficulties, such as appear in the damping treatment of Compton scattering, do not arise, and the " cut-off " of damping technique is not necessary. The same result applies to the relativistic collision of two electrons.

Damping Theory and the Propagation of Radiation, by J. HAMILTON.

ABSTRACT. It is shown that a careful examination of the concepts involved in radiation damping theory makes it possible to treat the propagation of energy from one atom to another through the electromagnetic field. An exact and continuous solution of the propagation problem is given.

Continuous γ-Emission in Neutron-Proton Collisions, by M. KROOK.

ABSTRACT. The cross-section for Bremstrahlung in neutron–proton collisions is calculated with a central exchange-force model and for energies up to 20 MEV. Its value is found to lie between 10^{-28} and 10^{-29} cm^2. This probably is too small to be measured with experimental techniques at present available.

Resonance Transfer of Electronic Energy in Organic Crystals, by E. J. BOWEN, E. MIKIEWICZ and F. W. SMITH.

ABSTRACT. The resonance transfer of energy in crystals of organic compounds when excited by visible or ultra-violet light has been examined by measuring the fluorescence of the systems naphthalene–anthracene, anthracene–naphthacene, naphthacene–pentacene, anthracene–acridine, anthracene–phenazine, and anthracene–di- and hexabromobenzenes. The conditions for "exciton " trapping and the relations of resonance transfer to fluorescence quenching are discussed.

Further Experiments with an Adjustable Geiger-Müller Counter, by A. G. FENTON and E. W. FULLER.

ABSTRACT. An adjustable Geiger-Müller counter and special electronic circuits have been used to study the effect of several counter variables on the occurrence of multiple pulses. The results show that the plateau characteristics combine the effects of an increase in efficiency and an increase in the probability of a multiple pulse as the operating potential is raised. It was found that the probability of occurrence of multiple pulses does not depend simply upon the charge per pulse, but that it depends also upon the anode diameter and the total pressure of the filling.

Curves have been drawn showing the charge per pulse as a function of the operating potential. A change of slope occurs in these curves when the charge per pulse is approximately equal to the charge on the active portion of the anode.

The counter was used as an ionization chamber at operating potentials below the Geiger region. It was found that as the potential is raised above the threshold of the proportional region the ionization current varies with potential as $\log I \propto V$ while at higher potentials the relation is adequately expressed by $\log \log I \propto V$. Near the Geiger region a sharp break occurs in the log log curve, which is taken to indicate that the discharge begins to spread at this stage.

The Steady Non-Uniform State for a Liquid, by R. Eisenschitz.

ABSTRACT. The Smoluchowski equation for the distribution of the relative coordinates of a representative pair of molecules in a liquid is formulated in such a way that the effect of non-uniform velocity or temperature is taken into account. The equation is reduced to an ordinary linear differential equation of the second order. One of the two constants of integration is readily determined whereas the determination of the other constant may lead to inconsistencies. The molecules are assumed to interact with the average potential that is appropriate to thermal equilibrium. If it is assumed that this potential becomes infinite of the first order at distance zero then the problem of viscous flow and the problem of thermal conduction are uniquely soluble. Approximate evaluation of the transport coefficients shows the right kind of temperature effect. The result is in accordance with the previous treatment by means of the cell model.

Dislocation Theory of Yielding and Strain Ageing of Iron, by A. H. Cottrell and B. A. Bilby.

ABSTRACT. A theory of yielding and strain ageing of iron, based on the segretation of carbon atoms to form atmospheres round dislocations, is developed. The form of an atmosphere is discussed and the force needed to release a dislocation from its atmosphere is roughly estimated and found to be reasonable. The dependence on temperature of the yield point is explained on the assumption that thermal fluctuations enable small dislocation loops to break away; these loops subsequently extend and cause yielding to develop catastrophically by helping other dislocations to break away. The predicted form of the relation between yield point and temperature agrees closely with experiment.

Strain ageing is interpreted as the migration of carbon atoms to free dislocations. The rate of ageing depends upon the concentration of carbon in solution and the estimated initial rate agrees with experiment on the assumption that about 0·003% by weight of carbon is present in solution.

Investigations on the Binding Energy of Heavy Nuclei, by R. Huby.

ABSTRACT. The binding energy of a heavy nucleus is one of the quantities which a satisfactory nuclear force theory ought to predict correctly. Approximate calculations of this binding energy are performed, taking as nuclear force the static interaction of Møller and Rosenfeld. The parameters appearing in the interaction have been fixed by data on light nuclei.

The Fermi-gas model is taken as basis for the calculations, a first solution being obtained by a first-order perturbation (or variation) method. For refinement, a second-order perturbation calculation is made, and Svartholm's variation-iteration method is attempted. The first-order results yield about 10% of the required energy, and the second-order about 40%; a reasonable prediction of the size of the nucleus is obtained. The reliability of the perturbation results is uncertain; Svartholm's method does not appear to be well suited to the investigation of heavy nuclei.

Two alternative modifications are made to the nuclear force to make it yield the correct binding energy according to the perturbation calculation.